

# Essays on Income Taxation and Labor Market Outcomes

Submitted by

Kazuhiko Sumiya

for the degree of Doctor of Philosophy

of the

Royal Holloway, University of London

## Abstract

This thesis addresses questions regarding the effect of income taxation on labor market outcomes.

Chapter 1 asks “Do income taxes create accumulating effects over time on pre-tax hourly wages by distorting on-the-job human capital accumulation?” This chapter presents micro evidence by exploiting Danish administrative data and a tax reform. After non-parametrically controlling for pre-reform income and covariates, I provide graphical evidence and difference-in-differences estimates regarding the accumulating effects. I find suggestive evidence that learning-by-doing is the underlying channel. However, I do not find significant effects of taxes on participation in on-the-job training courses. To understand welfare implications, this chapter next constructs a labor supply model with endogenous wage dynamics as a conceptual framework. Guided by the model, the elasticity estimates imply that the wage response has a larger impact on welfare than the hour response and that both responses create accumulating effects on welfare.

In Chapter 2, we study the impact of labor income taxation on workers’ job search behavior and the implications it has for the equilibrium allocation of heterogeneous workers across heterogeneous firms. The analysis is conducted within a complete markets equilibrium on-the-job search model with two-sided heterogeneity, endogenous job search effort and hiring intensity, equilibrium wage formation, and firm entry and exit. By appropriating part of the gain from finding a better paid job, income taxation reduces the return to job search effort, and distorts workers’ job search effort, which, in turn, distorts the equilibrium allocation of labor. The model is estimated on Danish matched employer-employee data, and is used to evaluate a series of tax reforms in Denmark in the 1990s and 2000s, to provide new insights into the elasticity of taxable labor income, and to identify a Pareto-optimal income tax reform.

# Contents

<b>List of Figures</b>	<b>6</b>
<b>List of Tables</b>	<b>7</b>
<b>1 Accumulating Effects of Income Taxes on Wages: Micro Evidence from Denmark</b>	<b>9</b>
1.1 Introduction . . . . .	9
1.2 Danish tax system and reform in 1987 . . . . .	12
1.3 Data . . . . .	14
1.4 Empirical strategy . . . . .	16
1.4.1 Treated and control units . . . . .	17
1.4.2 Endogeneity by pre-reform income . . . . .	18
1.4.3 Covariate balancing . . . . .	20
1.5 Difference-in-differences estimation . . . . .	22
1.5.1 Graphical evidence . . . . .	22
1.5.2 Threats to identification . . . . .	24
1.5.3 Regression analysis . . . . .	26
1.5.4 On-the-job training . . . . .	30
1.5.5 Learning-by-doing . . . . .	31
1.6 Welfare analysis . . . . .	32
1.6.1 Model . . . . .	33
1.6.2 Clarification . . . . .	35
1.6.3 Welfare implications . . . . .	36
1.7 Conclusion . . . . .	37
<b>2 Income Taxation and the Equilibrium Allocation of Labor</b>	<b>53</b>
2.1 Introduction . . . . .	53
2.2 The model . . . . .	56
2.2.1 The environment . . . . .	56

2.2.2	Household behavior . . . . .	60
2.2.2.1	The complete markets steady state labor allocation . . . .	64
2.2.2.2	Tax reforms and partial equilibrium substitution and in- come effects . . . . .	65
2.2.3	Firm behavior . . . . .	66
2.2.3.1	Profit maximization . . . . .	67
2.2.3.2	Optimal wage and hiring policies . . . . .	69
2.2.3.3	Firm entry and exit . . . . .	70
2.2.3.4	Aggregation . . . . .	71
2.2.4	Equilibrium . . . . .	72
2.3	Data . . . . .	73
2.3.1	Data sources . . . . .	73
2.3.2	Selecting the analysis data . . . . .	75
2.4	The empirical tax function . . . . .	76
2.4.1	The Danish income tax system 1990-2005 . . . . .	76
2.4.2	Estimating the marginal tax functions . . . . .	83
2.5	Model estimation . . . . .	85
2.5.1	Parameterization . . . . .	85
2.5.2	The government budget . . . . .	88
2.5.3	Auxiliary statistics . . . . .	88
2.5.4	Model fit . . . . .	93
2.5.5	Structural parameter estimates . . . . .	94
2.5.6	The equilibrium allocation of labor . . . . .	99
2.6	Equilibrium tax reform evaluations . . . . .	100
2.7	The long run elasticity of taxable labor income . . . . .	106
2.7.1	Decomposing the long run elasticity of taxable labor income . . . .	108
2.8	A Pareto optimal tax reform . . . . .	109
2.9	Conclusion . . . . .	113

## Appendices 114

A.1	Proofs and derivations . . . . .	115
A.1.1	Law of motions for $n_t^0(a)$ and $n_t^1(w a)$ . . . . .	115
A.1.2	Steady state labor allocation . . . . .	116
A.1.3	Applying the Maximum Principle . . . . .	118
A.1.4	Proof of Lemma 1 . . . . .	125
A.1.5	Proof of Lemma 2 . . . . .	125
A.1.6	Proof of Proposition 1 . . . . .	125

A.1.7	Proof of Proposition 2 . . . . .	125
A.1.8	Proof of Proposition 3 . . . . .	126
A.2	Details on the data . . . . .	127
A.2.1	The merging procedure . . . . .	127
A.3	Estimating $T'(\cdot)$ and $T(\cdot)$ . . . . .	130
A.4	Numerical implementation . . . . .	131
A.4.1	Solving the model . . . . .	131
A.4.1.1	Discretization of $H(a)$ and $\Gamma_0(p)$ . . . . .	131
A.4.1.2	Finding $\lambda(a)$ , $v(a p)$ , and $w(a p)$ . . . . .	131
A.4.1.3	Calculate $w(a p)$ . . . . .	133
A.4.1.4	Calculate $v(a p)$ . . . . .	133
A.4.1.5	Calculate $\lambda(a)$ . . . . .	133
A.5	Allocative distortions and tax progressivity . . . . .	134
<b>Bibliography</b>		<b>135</b>

# List of Figures

1.1	Overview of the 1987 tax reform . . . . .	41
1.2	Correlation between treatment assignment and $\log LI_{i86}$ . . . . .	42
1.3	Graphical evidence . . . . .	43
1.4	Robustness of MB group . . . . .	44
1.5	Actual tax brackets of MB group . . . . .	45
1.6	Density around the middle bracket cutoff . . . . .	46
1.7	Composition change in employed workers . . . . .	47
1.8	Wage response . . . . .	48
1.9	Exclusion restriction test . . . . .	49
1.10	On-the-job training response . . . . .	50
1.11	Hour response . . . . .	51
1.12	Welfare implication of wage and hour responses . . . . .	52
2.1	Average personal income, capital income, and deductions by labor income percentiles . . . . .	80
2.2	Marginal tax function and tax function for Denmark 1999-2003 . . . . .	85
2.3	Job destructions and match output . . . . .	96
2.4	Key equilibrium objects . . . . .	97
2.5	Estimated annual job offer arrival rates by worker- and firm-type . . . . .	99
2.6	The equilibrium allocation of labor . . . . .	100
2.7	Danish tax regimes 1990-2005: Marginal and average tax functions . . . . .	101
2.8	A Pareto optimal tax reform . . . . .	111
E.1	Counterfactual linear tax systems . . . . .	134

# List of Tables

1.1	Income concept in the Danish tax system . . . . .	38
1.2	Danish tax system before and after the 1987 reform . . . . .	38
1.3	Descriptive statistics in 1986 . . . . .	38
1.4	Pre-treatment covarites (MB group in 1986) . . . . .	39
1.5	Pre-reform 1986 covarites (employed in 1993) . . . . .	39
1.6	Pre-treatment covarites (MB group in 1986) . . . . .	40
2.1	November 28th cross-sections in the 1999-2003 estimation panel . . . . .	76
2.2	Summery Statistics for the 1999-2003 estimation panel . . . . .	77
2.3	Income concepts in the Danish income tax system . . . . .	78
2.4	The summary statistics on the distribution of 1999-2003 taxable income components . . . . .	79
2.5	Danish income tax bases and rates 1990-2005 . . . . .	82
2.6	Model fit . . . . .	95
2.7	Structural parameter estimates . . . . .	95
2.8	Evaluating the 1990, 1994, and 2004 Danish income tax reforms for nonem- ployment, labor income and the government budget . . . . .	103
2.9	Long run taxable labor income elasticities . . . . .	107
2.10	Decomposing the full equilibrium elasticity of taxable labor income . . . . .	109
2.11	Percentage changes between 1999-2003 regime and optimally reformed regime— Partial Equilibrium . . . . .	111
2.12	Percentage changes between 1999-2003 regime and optimally reformed regime— Full Equilibrium . . . . .	112
E.1	The impact of income tax progressivity on nonemployment, labor income and the government budget . . . . .	136

## Acknowledgements

I would like to express my gratitude to a number of people at Royal Holloway who have encouraged and supported me in completing this thesis. I would like to thank my supervisor, Jesper Bagger, for all his advice and guidance. I am also deeply grateful to Manolis Galenianos, Ahu Gemici, Melanie Lührmann, Juan Pablo Rud, and Ija Trapeznikova for helpful comments and kind words of encouragement.

To a number of people at Aarhus University. I am especially grateful to Henning Bunzel, Bent Jesper Christensen, Mads Hejlesen, and Rune Vejlin for their support.

To Jesper Bagger, Mads Hejlesen, and Rune Vejlin for allowing me to include our paper in this thesis.

To all my fellow PhD students in London and Aarhus.

And finally, to my family for their enduring faith in me.



# Chapter 1

## Accumulating Effects of Income Taxes on Wages: Micro Evidence from Denmark

### 1.1 Introduction

Economists have paid a large amount of attention to distortionary effects of income taxes on individual behavior. The early empirical literature exclusively focuses on a labor supply response and reaches its consensus that taxes have an impact almost only on labor force participation by females.<sup>1</sup> However, observing that taxes distort not only labor supply but also behavior in other dimensions (e.g., work effort or tax avoidance), Feldstein (1995, 1999) argues that an elasticity of taxable income (ETI hereafter) with respect to marginal tax rates is more appropriate for welfare analysis because it works as a sufficient statistic capturing all the behavioral responses. Following this contribution, the recent literature shifts its interest from the elasticity of labor supply to the ETI as a relevant parameter.<sup>2</sup>

Recent estimates of the ETI using US data are around 0.4 (Saez et al. (2012b)), but heterogeneity in two aspects is worth mentioning. The first is along the income distribution. Gruber and Saez (2002) find larger ETIs for high-income individuals than for the rest. The second is that ETIs for self-employed workers are larger than for wage earners as found by Kleven and Schultz (2014a) using Danish data. The literature attributes these

---

<sup>1</sup>See Blundell and MaCurdy (1999) and Blundell, MaCurdy, and Meghir (2007) for surveys on theoretical and empirical frameworks. Keane (2011) and Meghir and Phillips (2010) give overviews of recent empirical evidence.

<sup>2</sup>See Chetty (2009a,b), Hendren (2016), and Kleven (2018) for the “sufficient statistic” approach, and see Saez, Slemrod, and Giertz (2012b) for a survey on the ETIs. Slemrod and Kopczuk (2002) formulate the idea that the ETI is not an immutable structural parameter but a choice variable shaped by features of the tax system, e.g., opportunities for tax avoidance or evasion.

results to tax avoidance and evasion.<sup>3</sup> In other words, ETIs are small for the majority of employed workers, who consist of more than half of the entire labor force in most countries (e.g., Kleven (2014)). Slemrod (1992, 1995) argues that real behavior is far less responsive to taxes than avoidance or timing behavior. His argument implies that small ETIs for employed workers reflect real behavioral responses such as labor supply, which are crucial for economic performance and growth.<sup>4</sup> To better understand distortionary effects of taxes, one would need to move beyond the ETI because its aggregate notion might miss some important micro responses.

This paper empirically investigates a new channel through which taxes distort real behavior among employed workers. More specifically, the hypothesis is that taxes have an impact on pre-tax hourly wages. Despite the fact that wages are an important labor market variable along with labor supply variables, their responsiveness to taxes has not been studied in detail. Related research is done by Blomquist and Selin (2010), and I compare my estimation results to theirs in Section 1.5.<sup>5</sup>

Taxes can have an impact on wages through several behavioral responses. The empirical literature has analyzed the effect of taxes on job search (Chetty, Friedman, Olsen, and Pistaferri (2011b) and Gentry and Hubbard (2004b)) or bargaining (Piketty, Saez, and Stantcheva (2014)). Although they do not look at pre-tax hourly wages as an outcome, both responses arguably have an impact on wages. This paper contributes to the literature by presenting empirical evidence on another behavioral response to taxes: human capital accumulation on the job. This channel has been studied mainly by structurally estimating Ben-Porath models (Guvenen, Kuruscu, and Ozkan (2014), Heckman, Lochner, and Taber (1998, 1999), and Taber (2002)) or learning-by-doing models (Keane (2015, 2016) and Keane and Wasii (2016)). Rather than taking a structural approach, this paper estimates reduced-form parameters, which will provide a basis for structural estimation. Since on-the-job human capital accumulation can be through on-the-job training (e.g., Simonsen and Skipper (2008)) or learning-by-doing (e.g., Stinebrickner, Stinebrickner, and Sullivan (2017)), I distinguish these two by using information on participation in on-the-job training courses and hours worked as proxies.<sup>6</sup>

---

<sup>3</sup>For example, Goolsbee (2000) and Gorry, Hassett, Hubbard, and Mathur (2017) find that large ETIs among executives come from the exercise of stock options as a way to shift income intertemporally, although Hall and Liebman (2000) find only a modest impact of stock options. Kreiner, Leth-Petersen, and Skov (2016) also identify large intertemporal income shifting driven by a tax reform among high-income employees.

<sup>4</sup>See Mertens and Olea (2018) for recent time-series evidence regarding impacts of taxes on GDP and unemployment.

<sup>5</sup>Lockwood and Manning (1993) and Lockwood, Slok, and Tranaes (2000) are early examples using aggregate data. Although their focus is not on individual income taxes, Saez, Schoefer, and Seim (2017) find zero effects of an employer payroll tax cut on net-of-tax wages while Fuest, Peichl, and Sieglöcher (2018) and Suárez Serrato and Zidar (2016) estimate the incidence of corporate taxes on wages.

<sup>6</sup>Education choice (i.e., human capital investment off the job) is also an important behavioral response

This paper also makes a contribution by considering accumulating effects of taxes on wages. As a worker invests in human capital subject to taxes for multiple periods, distortion will accumulate over time and appear gradually in her wage dynamics. One thus has to analyze not only a one-shot impact but also a long-lasting impact of taxes for an appropriate policy evaluation. However, broadly speaking, the empirical micro literature estimates short-run static elasticities, while the macro calibration literature compares two long-run steady-states.<sup>7</sup> The current paper therefore fills the gap between these two strands of the literature.

To find micro evidence, this paper exploits rich Danish administrative data covering the whole population with detailed individual income and demographics. Exogenous variation in tax rates comes from a large tax reform that pushes some taxpayers to other tax brackets. Comparing taxpayers pushed to another bracket with those staying in the same bracket constitutes a difference-in-differences (DID hereafter) empirical strategy of this paper. To deal with endogeneity caused by correlation between treatment assignment and pre-reform income, I focus on a specific tax bracket where treatment and control groups have almost identical distributions of pre-reform income, leading to non-parametric controlling for it. In addition, to make the treatment group similar to the control group in terms of pre-reform covariates, I apply a covariate balancing method akin to propensity score matching methods.

I first provide graphical evidence that demonstrates parallel wage dynamics between the treatment and control groups before the reform but diverging wage dynamics after the reform. Moving on to regression analysis, DID estimates show that an intertemporal elasticity of wages with respect to net-of-marginal tax rates (i.e., one minus marginal tax rates) is statistically significant and 0.05 one year after the reform but reaches as high as 0.1 five years after the reform. Several threats to identification such as composition changes in employed workers over time are discussed and tested to verify my estimation results. Further, I find suggestive evidence that learning-by-doing is the underlying channel. Namely, an intertemporal elasticity of hours worked with respect to net-of-marginal tax rates is increasing over time. The increasing patterns of both wage and hour elasticities over time are consistent with the notion of learning-by-doing. However, I do not find significant effects of taxes on the incidence of on-the-job training.

Since the literature typically analyzes a welfare consequence of income taxes under a static labor supply model with exogenous wages, this paper next studies how the accumulating effects of taxes on wages change our understanding on welfare. Although the

---

but identification will be extremely difficult.

<sup>7</sup>A few exceptions in the ETI literature estimating long-run elasticities are Holmlund and Soderstrom (2011), who adopt a dynamic panel model, and Giertz (2010) and Heim (2009), who consider time-lagged responses.

ETI is a sufficient statistic for welfare analysis and one thus does not have to identify each underlying response (such as wage and hour responses), decomposing welfare into its roots (i.e., “the anatomy of behavioral response”) opens up the black box and has important normative implications (Chetty et al. (2011b), Saez (2003), and Slemrod (1996, 1998)).

As a conceptual framework, I construct a labor supply model in which a worker chooses continuous hours to work and invests in on-the-job training subject to taxes every period. Wages increase through learning-by-doing or on-the-job training. I then derive a formula for marginal excess burden of taxation, which is decomposed into wage and hour elasticities. Guided by this model, the elasticity estimates imply that the wage response has a larger impact on welfare than the hour response and that both responses create accumulating effects on welfare. In relation to “the anatomy of behavioral response”, these results indicate that when evaluating tax policies aimed at encouraging labor supply such as Earned Income Tax Credit (EITC), researchers should look at responses not only in labor supply but also in wages and their interactions over time. Further normative implications will be negative impacts on labor supply and welfare created by phase-out (positive) tax rates of EITC could be mitigated by other tax policies encouraging human capital accumulation on the job.

The paper proceeds as follows: Section 1.2 describes a tax system and reform in Denmark. Section 1.3 explains data. Section 1.4 explains empirical strategies. Section 1.5 presents DID estimation results. Section 1.6 analyzes welfare. Section 1.7 concludes. All the tables and figures are placed at the end of the paper.

## **1.2 Danish tax system and reform in 1987**

Like other Scandinavian countries, Denmark is characterized by relatively high tax burdens. According to Kleven (2014), the ratio of tax revenue to GDP is 48 percent in 2012. This is higher than other countries such as Germany (36 percent), the United Kingdom (35 percent), and the United States (25 percent). Denmark collects about half of its revenue from individual income taxes, and here their impact on individual behavior is studied by using a tax reform as a natural experiment.

The Danish tax system has experienced several reforms since the 1980s, and major reforms took places in 1987, 1994, 1999, and 2004. I focus on the 1987 reform because it creates larger variation in tax rates, tax bases, and tax bracket cutoffs. In addition, compared to other reforms, the tax system is stable before and after the 1987 reform. This reform is therefore not gradual changes phased in over an extended period of time but a one-shot large change taking place in 1987, which is suitable for estimating accumulating effects of taxes on wages.

Before the 1987 reform, individual income taxation in Denmark was based on a single measure of taxable income. Table 1.1 explains the components of taxable income and their main items.<sup>8</sup> Taxable income (TI) is composed of personal income (PI), capital income (CI), and net of deductions ( $-D$ ). Personal income is further decomposed into labor income (LI) and other personal income (OPI). Labor income is the largest component of taxable income as confirmed by descriptive statistics in Section 1.3, and the interest of this paper lies on marginal tax rates on labor income.

The left panel of Table 1.2 describes the individual tax system before the 1987 reform. First of all, one can see that the tax base is simply taxable income (TI) in all brackets. Taxes in Denmark are divided into regional and national taxes. Although their variations are small, regional tax rates vary by municipality and county.<sup>9</sup> National taxes have a progressive structure with three brackets. These tax rates are cumulative such that, for example, an individual in the middle bracket faces a  $28.0 + 19.9 + 14.4 = 62.3$  percent marginal tax rate. Finally, while taxation is based on individual filing for married couples, some exemptions can be transferred across spouses. As I explain in Section 1.4, these details of the tax system are taken into account in simulating tax liabilities and effective marginal tax rates for individuals.

Given that the reform changed the tax system in a slightly complicated way, it will be worthwhile to get its big picture. Figure 1.1 plots marginal tax rates on labor income (LI) as a function of LI before and after the 1987 reform. The tax rates and cutoffs are from Table 1.2 while assuming other personal income (OPI), capital income (CI), and deductions (D) are all zero. By assuming  $OPI = CI = D = 0$ , marginal tax rates are a deterministic function of only LI, which simplifies a complex structure of the tax system and thus is helpful for a overview of the reform. All the details are explained below but it is clear from the figure that, for example, the reform makes the tax system less progressive.

Before moving on to details of the reform, it is also useful to mention its background. In an international context, the Danish tax system prior to the reform was characterized by high marginal tax rates and narrow tax bases. As the left panel of Table 1.2 shows, the top marginal tax rate was as high as 73 percent. Regarding narrow tax bases, first note that capital income is negative for the majority of Danish taxpayers as a result of interest payments on debt such as mortgage and other loans. Therefore, negative capital income was subtracted from the tax base of every bracket, leading to narrow tax bases and creating incentives to getting into debt. Given these backgrounds, the purpose of the reform was, among others, to lower marginal tax rates and encourage private savings. The reform was passed in parliament in March 1986 and came into effect from January 1987.

---

<sup>8</sup>Description about institutional settings is partly based on Kleven and Schultz (2014a), who exploit a series of tax reforms in Denmark including the 1987 one.

<sup>9</sup>In 1986, for example, the 10th percentile of regional tax rates is 26.1 while the 90th percentile is 29.8.

The right panel of Table 1.2 describes the individual tax system after the 1987 reform. As one can see, the reform is characterized by (i) a differential change in progressive tax rates across brackets, (ii) a change in tax bases from a single measure of taxable income (TI) to its components (PI, CI, D) for the middle and top brackets, and (iii) an increase in bracket cutoffs in national taxes. Regarding the third point, the cutoffs increased by around 15 percent for the bottom and middle brackets and by 7.5 percent for the top bracket. Note that the inflation rate in 1986 was 3.7 percent, and thus the bracket cutoffs increased by larger amounts than inflation indexation. Given the second and third points that the tax bases and cutoffs have changed by the reform, some taxpayers were pushed upward to the higher bracket while others with the same taxable income (TI) were pushed downward to the lower bracket depending on the composition of TI. This bracket movement gives me a DID identification strategy that compares workers staying in the same bracket with those pushed to another bracket. I explain the details of this empirical strategy in Section 1.4.

As is also emphasized by Kleven and Schultz (2014a), it is important to note that the reform changed the tax structure along all income groups. If the reform changed the tax structure only among top-bracket taxpayers, comparing them with middle-bracket taxpayers (as a control group) over time would be conflated by non-tax factors that create heterogenous wage trends such as skill-biased technological changes. This reform, however, affected every taxpayer in a heterogenous way depending on the composition of pre-reform taxable income. Therefore, one can find two groups of taxpayers with similar pre-reform income levels but with different tax-bracket movements. As I show in Section 1.4, this feature of the reform enables me to find treatment and control groups whose pre-reform income distributions are almost identical to each other.

Additional comments are in order about the reform. First, regional taxes changed only marginally throughout the reform. The variation thus comes almost exclusively from national taxes. Second, as is consistent with its background, the reform broadened the tax bases of the middle and top brackets because negative capital income and deductions cannot be subtracted from these tax bases under the new tax system.

### 1.3 Data

The empirical analysis of this paper exploits an administrative data set based on several registers such as the tax register and IDA covering the whole Danish population since 1980.<sup>10</sup> They are constructed and maintained by Statistics Denmark for research pur-

---

<sup>10</sup>IDA is an acronym for “Integreret Database for Arbejdsmarkedsforskning” (Integrated Database for Labor Market Research).

poses. The data set contains a wide range of information including worker identifiers, socioeconomic backgrounds, job characteristics, and individual income necessary to simulate tax liabilities and effective marginal tax rates. Individual income is observed yearly and aggregated into labor income (LI), other personal income (OPI), capital income (CI), and deductions (D) as listed in Table 1.1.

For socioeconomic backgrounds, I use yearly variables on age, gender, marital status, the number of children (younger than 17 years old), education levels (low, middle, and high)<sup>11</sup>, and household assets. Let me clarify a few points regarding household assets. First, they include financial assets, non-financial assets, and (net of) debts at their market prices. Second, they are summation of an individual’s and his or her spouse’s assets. This wide range of demographic information allows me to control for factors related to wage trends and will be one of the advantages of my data set given that tax return data in US usually does not have such information as pointed out by Weber (2014).

For an outcome variable, I first use hourly wages as a main focus of this paper. We observe in the data whether workers are employed or not on the 28th of November each year. The definition of employment in this paper is hence being employed on this particular day. Then, only for employed workers, the data contains hourly wages for the jobs held on the 28th of November. For simplicity, a job held on the 28th of November is referred to as a “November job” in the following. Hourly wages are computed by dividing yearly earnings from November jobs (precisely recorded by the tax authorities) by yearly hours worked for November jobs.<sup>12</sup> Hours worked are estimates based on yearly pension contribution records (called “ATP”) by exploiting the fact that accumulated pensions depend only on hours in a certain way. The details of this estimation are documented thoroughly by Lund and Vejlin (2016), who find that the estimates are precise especially for full-time workers by comparing their wage estimates to wages obtained from another data source.

I next use daily hours worked as an outcome. Given the estimates of yearly hours worked for November jobs, daily hours worked are computed by dividing yearly hours worked by yearly days worked for November jobs. The latter information comes from job spell data that contains start and end dates of the November jobs. Since the job spell data starts in 1985, daily hours worked are available only from 1985. Although they are estimates, availability of both wages and hours together with the tax return data is novel and allows one to analyze responses to taxes in both dimensions. Finally, I also use information on participation in on-the-job training courses as an outcome but explain its

---

<sup>11</sup>Low education is defined by completion of primary education. Middle education is defined by completion of high school or vocational education. High education is defined by holding of bachelor, master, or PhD degrees.

<sup>12</sup>Labor income (LI) in Table 1.1 includes earnings from a November job and jobs held outside of the 28th of November. That is, LI is the yearly total amount of earnings.

details in the corresponding analysis of Section 1.5.

A window of analysis is between 1983 and 1993 because there were minor tax reforms in 1982 and 1983, and also because the next tax reform (after the 1987 one) took place in 1994. I select males employed (on the 28th of November) in all of the pre-reform years from 1983 to 1986. This criteria picks up male workers who are strongly attached to the labor market and are thus presumably in full-time jobs, which is desirable for the following reasons: First, as Lund and Vejlin (2016) find it, the wage estimates of full-time workers are more precise than those of part-time workers. Second, since I test the hypothesis that taxes distort individual behavior on the job, it is necessary to select core labor market participants. As Chetty et al. (2011b) find it, (married) female workers form bunching around bracket cutoffs in Denmark. Their behavior might be described by avoidance or timing responses in addition to real behavioral responses such as human capital accumulation. It thus will be difficult to estimate accumulating effects of taxes on wages with females. I come back to the issue of bunching in Section 1.5.

This sample selection also ensures wage observations for all of the pre-reform years and therefore is useful for looking at pre-reform wage trends. Notice that the restriction is imposed only on the pre-reform period because, after the reform, a decision on taking a job will be endogenous in relation to the reform. This point is related to composition changes in employed workers over time, which I discuss in Section 1.5.

In the end, my data set consists of male workers employed in the pre-reform years and follows them after the reform with small attrition due to, e.g., death or emigration (around 3 percent attrition in the final year of the sample period). Table 1.3 displays cross-sectional descriptive statistics of these workers in 1986. Mean values as of 1986 are listed, and 1 Danish Krone (DKK) in 1986 is approximately equal to 0.3 US Dollar (USD) in 2016. A few points are worth mentioning about the components of taxable income. First, labor income is by far the largest component, suggesting that it is important to control for it in the analysis to come. Second, as I touched upon in Section 1.2, capital income is, on average, negative mainly as a result of interest payments on debt such as mortgage and other loans.<sup>13</sup>

## 1.4 Empirical strategy

This paper adopts a DID framework by using the tax reform and administrative data described in Sections 1.2 and 1.3. The basic idea of DID estimation is comparing outcomes of treated and control units over time where treated units get some treatment after an exogenous event, which is the tax reform in this paper. The key identification assumption

---

<sup>13</sup>As the data includes only males, I use masculine pronouns for a worker below.



is a parallel-trend assumption, which requires that treated units have the same trend in outcomes as control units in the absence of the treatment. Although this assumption is not directly testable, one could conduct a (placebo) test to see whether the pre-treatment trends are identical between the two units.

In the following, I explain (i) how to form treated and control units while exploiting the reform, (ii) how to deal with endogeneity caused by correlation between treatment assignment and pre-reform income, and (iii) a covariate balancing method that will be useful for the parallel-trend assumption.

### 1.4.1 Treated and control units

As I mentioned in Section 1.2, this paper exploits the variation in tax bases and bracket cutoffs induced by the reform that pushes some taxpayers to other tax brackets. For this purpose, it's necessary to identify in which brackets taxpayers are located but the tax register does not contain such information. I hence calculate yearly tax liabilities for individuals using a tax simulator based on the one developed by Kleven and Schultz (2014a)<sup>14</sup>. This simulator requires, as inputs, individual income listed in Table 1.1 (i.e., LI, OPI, CI, and D) and individual characteristics such as marital status (to calculate exemption transfers across spouses). The simulator takes into account the details of the Danish tax system and pins down in which brackets taxpayers are located.

Tax-liable individuals are located in one of the three brackets as shown in Table 1.2. Let  $B^{86}(z_{i86})$  denote that a worker  $i$  is in the bottom bracket under the 1986 tax system with his 1986 income and characteristics  $z_{i86} = \{LI_{i86}, OPI_{i86}, CI_{i86}, D_{i86}, x_{i86}\}$ , where  $x_{i86}$  includes, for example, marital status. Similarly, one can define  $M^{86}(z_{i86})$  and  $T^{86}(z_{i86})$  as a worker  $i$  in the middle or top brackets under the 1986 tax system with his 1986 income and characteristics  $z_{i86}$ .

Let us next construct a measure of bracket movements created by the 1987 reform. This measure should be mechanical in the sense that it captures exogenous variation induced by the reform. To this end, it is useful to consider the following counterfactual bracket location:  $\tilde{B}^{87}(z_{i86})$  denotes that a worker  $i$  is in the bottom bracket with his 1986 income and characteristics  $z_{i86}$  if the tax system were the inflation-adjusted 1987 tax system. By the inflation-adjusted 1987 tax system, I mean the 1987 tax system where all the monetary values such as bracket cutoffs are deflated at the 1986 price level.<sup>15</sup> Note that his behavior and income are fixed at the 1986 level  $z_{i86}$ .  $\tilde{M}^{87}(z_{i86})$  and  $\tilde{T}^{87}(z_{i86})$  are analogously defined.

By combining the actual and counterfactual bracket locations, I hence construct a

<sup>14</sup>I modified their codes available on the website of American Economic Association.

<sup>15</sup>I use CPI downloaded from the website of Denmark Statistics as a deflator.

measure of bracket movements, that is, mechanical changes in brackets as follows:

$$\begin{aligned}
& B^{86}(z_{i86})\tilde{B}^{87}(z_{i86}) : i \text{ stays in the bottom bracket (BB)} \\
& B^{86}(z_{i86})\tilde{M}^{87}(z_{i86}) : i \text{ moves from the bottom to middle brackets (BM)} \\
& M^{86}(z_{i86})\tilde{B}^{87}(z_{i86}) : i \text{ moves from the middle to bottom brackets (MB)} \\
& \quad \vdots \\
& T^{86}(z_{i86})\tilde{T}^{87}(z_{i86}) : i \text{ stays in the top bracket (TT)}.
\end{aligned}$$

To ease the notations, let BM denote  $B^{86}(z_{i86})\tilde{M}^{87}(z_{i86})$  with the same rule applied to other cases. For example, BM means that, when fixing his behavior and income at the 1986 level  $z_{i86}$ , he is in the bottom bracket under the 1986 tax system but in the middle bracket under the 1987 tax system. Therefore, BM is a movement from the bottom bracket to the middle bracket mechanically created by the reform. Rich and heterogenous variation in tax bases and bracket cutoffs induced by the reform generates all the movements of (BB, BM, MB, MM, MT, TM, TT).

Using the mechanical changes in brackets, the following four groups each with treated and control units are formed:

$$\begin{aligned}
& \text{BM group} = \{\text{Treated: BM, Control: BB}\} \\
& \text{MB group} = \{\text{Treated: MB, Control: MM}\} \\
& \text{MT group} = \{\text{Treated: MT, Control: MM}\} \\
& \text{TM group} = \{\text{Treated: TM, Control: TT}\}.
\end{aligned}$$

Treated units are pushed upward or downward to other brackets while control units stay in the same bracket. Comparing wage dynamics of treated units to those of control units before and after the reform constitutes the DID empirical strategy of this paper, as is similarly done by Saez (2003) and Singleton (2011), who also exploit bracket movements to estimate ETIs. I exclusively focus on the MB group for the identification reason discussed in Section 1.4.2 and test my hypothesis that the treated units have higher wage growth than the control units in the MB group.

### 1.4.2 Endogeneity by pre-reform income

As is clear from the discussion in Section 1.4.1, treatment assignment is a deterministic function of pre-reform income and characteristics  $z_{i86} = \{LI_{i86}, OPI_{i86}, CI_{i86}, D_{i86}, x_{i86}\}$ , where  $x_{i86}$  includes, for example, marital status. While it is straightforward to control for

$x_{i86}$  such as marital status, the ETI literature finds that estimates are very sensitive to specifications of pre-reform income controls. Since similar concerns will also apply to the relationship between taxes and wages, I here explain how to deal with endogeneity caused by correlation between treatment assignment and pre-reform income. My identification strategy is, in a nutshell, to control for labor income  $LI_{i86}$  carefully but leave the other components  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  uncontrolled and use them for variation.

Regarding pre-reform labor income  $LI_{i86}$ , it is easy to understand the endogeneity issue by starting from Figure 1.2, which plots kernel density estimates of the logarithm of  $LI_{i86}$  ( $\log LI_{i86}$ ) by treatment status for the BM, MB, MT, and TM groups. One can see that treatment assignment is correlated with  $LI_{i86}$  in an expected way: that is, in the BM and MT groups, the treated units have higher  $LI_{i86}$  (i.e., located at the right end of the brackets), while in the MB and TM groups, they have lower  $LI_{i86}$  (i.e., located at the left end of the brackets). This causes endogeneity because  $LI_{i86}$  is arguably also correlated with wage dynamics due to, e.g., skill-biased technological changes.

One needs to control for pre-reform income to deal with the endogeneity but the literature finds that regression results are very sensitive to specifications of controls.<sup>16</sup> This is easily understood from Figure 1.2. Let us consider the TM group for illustration. Since the support of  $\log LI_{i86}$  does not sufficiently overlap with each other between the treated and control units, linear regression with pre-reform income controls heavily relies on extrapolation.<sup>17</sup> However, the MB group does not suffer from lack of the common support. Rather, as Figure 1.2 clearly shows, the distributions of the treated and control units are almost identical to each other, which will ensure estimation results robust to specifications of  $LI_{i86}$  controls. I hence exclusively focus on the MB group in the following analysis.

Let us move on to the other components of pre-reform income  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$ . Since treatment assignment is a deterministic function of pre-reform income, controlling for  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  in addition to  $LI_{i86}$  will destroy identification. Given that labor income is the dominant component of taxable income (Table 1.3) and that the wide range of demographic information is available for controls, I do not control for the other components of pre-reform income to maintain the variation in treatment assignment. The identification assumption imposed here is therefore that once controlling for  $LI_{i86}$  and demographics (age, marital status, the number of children, education levels, and household assets),  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  do not affect wage dynamics. Borrowing terminology from the instrumental variable estimation literature, I call this assumption as an exclusion

<sup>16</sup>See Saez et al. (2012b) for extensive discussions, and see Burns and Ziliak (2017) and Weber (2014) for recent developments.

<sup>17</sup>Abadie, Diamond, and Hainmueller (2015) and Imbens (2015) clarify this point in the context of the advantage of synthetic control and matching methods over regression.

restriction:

$$\text{wage dynamics} \perp \{\text{OPI}_{i86}, \text{CI}_{i86}, \text{D}_{i86}\} \mid \{\text{LI}_{i86}, \text{demographics}\}, \quad (1.1)$$

which is tested and verified later in Section 1.5. As I explained in Section 1.2, the reform changed tax bases as well as bracket cutoffs and thus creates the variation in treatment assignment among those who have same  $\text{LI}_{i86}$  but different  $\{\text{OPI}_{i86}, \text{CI}_{i86}, \text{D}_{i86}\}$ .

### 1.4.3 Covariate balancing

In the DID framework, similarity between the treated and control units is crucial for the parallel-trend assumption. If they are different in terms of observable covariates, one cannot expect that their outcomes follow the same trends in the absence of the treatment. I thus apply a covariate balancing method to pre-treatment covariates of the MB group. The idea of covariate balancing methods is to construct weights to achieve balance in covariates between the treated and control units. The covariates considered here are (i) demographics in 1986 (age, age square, marital status, the number of children, education levels, and household assets), (ii) real wage changes between 1985 and 1986 ( $\Delta \log w = \log w_{i86} - \log w_{i85}$ ), and (iii) logarithms of labor income in 1986 ( $\log \text{LI}$ ) and its square. These are listed in Table 1.4, and I clarify a few points about the second and third covariates.<sup>18</sup>

The motivation to control for  $\Delta \log w$  is to deal with mean reversion. Recall that workers are employed in both of 1985 and 1986 by sample selection and located in the middle bracket in 1986 by construction of the MB group. Those hit by high transitory shocks in wages and labor income in 1986 are more likely to be located at the right end of the bracket and therefore be in the control units. This is shown by  $\Delta \log w$  and  $\log \text{LI}$  in Table 1.4. Due to mean reversion, their wages will drop on average for the following periods even without any tax reforms, causing overestimation of treatment effects. “Ashenfelter’s dip” found by Ashenfelter (1978) is an early example related to mean reversion in the DID framework, and the recent ETI literature proposes several solutions. Auten and Carroll (1999) control for log pre-reform income while Gruber and Saez (2002) propose more flexible specifications by using ten-piece splines in log pre-reform income. Kopczuk (2005) controls for log pre-reform income in levels and changes. My specification is similar to Kopczuk (2005) in the sense that I control for pre-reform income ( $\log \text{LI}$ ) and pre-reform wage changes ( $\Delta \log w$ ). Although these specifications are widely accepted in the ETI literature, Weber (2014) points out that their validity depends on assumptions regarding

---

<sup>18</sup>I use household assets not in logarithms but in levels because some households have negative assets due to debts.

serial correlation in error terms. If, for example, the error process is serially correlated, she shows that endogeneity remains unresolved. Therefore, it will be useful to test serial correlation to further examine this concern.

Regarding log LI, I experiment with the following two specifications: (i) I do not include any LI controls in the covariates and (ii) I include log LI and its square. Since the MB group has the almost identical pre-reform LI distributions between the treated and control units, these two specifications should produce similar estimation results. Therefore, it will be interesting to compare results with and without LI controls.

Returning to a covariate balancing method, I use entropy balancing developed by Hainmueller (2012). The popular alternative will be matching methods based on propensity scores, which are known to be sensitive to model miss-specifications. Entropy balancing however directly calculates weights of control units for pre-specified covariates and is thus non-parametric because there is no need to use parametric propensity scores.<sup>19</sup>

Columns 1 and 2 of Table 1.4 display raw mean values of the covariates by treatment status. The two units are different in several dimensions. In particular, the treated units have a much higher marriage rate. This is partly because some rules applied to the middle tax bracket about exemption transfers across spouses have changed after the reform. On the other hand, as expected, log LI is almost same between the two units with respect to the first and second moments. Column 3 displays mean values of the control units weighted by entropy balancing when I do not include any LI controls in the covariates. The results clearly show that entropy balancing works well to achieve high balance, which ensures that the pre-treatment covariates are non-parametrically controlled for. Column 4 displays the results when I include log LI and its square. Although I include only the first and second moments of log LI, the third and fourth moments are also found to be balanced.<sup>20</sup> Balancing log LI and its square is hence sufficient. In the following analysis, the control units are weighted (or balanced) ones unless explicitly stated otherwise. Combination of the DID design with the covariate balancing method will be useful to deliver causal effects, as is done in the literature based on matching on propensity scores (Abadie (2005), Blundell, Costa Dias, Meghir, and van Reenen (2004), Heckman, Ichimura, and Todd (1997), and Smith and Todd (2005)).

---

<sup>19</sup>See Hainmueller and Xu (2013) for implementation in Stata, Zhao and Percival (2017) for theoretical properties related to doubly robustness, and Marcus (2013) and Marcus and Siedler (2015) for economic applications.

<sup>20</sup>Mean of  $(\log LI)^3$  is 1709 for the both units. Mean of  $(\log LI)^4$  is 20445 also for the both.

## 1.5 Difference-in-differences estimation

With all the ingredients necessary for DID estimation having been explained in the previous sections, we are now in a position to look at accumulating effects of taxes on wages. I first present graphical evidence followed by a discussion on threats to identification. I next, one by one, present regression results with wages, participation in training courses on the job, and daily hours worked as outcomes.

### 1.5.1 Graphical evidence

Figure 1.3 plots mean log wages of the treated and control units for the MB group. Mean log wages are relative to 1986 (the final year of the old tax system), that is,  $\overline{\log w_{it}} - \overline{\log w_{i86}}$  is plotted for  $t = 83, \dots, 93$ , where  $w_{it}$  is real hourly wages for jobs held on the 28th of November in a year  $t$ . The control units are weighted by entropy balancing. The top panel of Figure 1.3 does not include any LI controls in the covariates to balance, which corresponds to Column 3 of Table 1.4. On the other hand, the bottom panel includes log LI and its square, which corresponds to Column 4 of Table 1.4.

First of all, the two panels of Figure 1.3 show quite similar wage dynamics of the control units.<sup>21</sup> The ETI literature finds estimates very sensitive to specifications of pre-reform income controls. For example, Saez et al. (2012b) find that, using the tax reform targeted at the top income group, some signs of ETIs are flipped once controlling for pre-reform income. This is partly because one has to compare the top (e.g., 1%) income group with the next (e.g., 9%) income group and thus heavily rely on extrapolation in regression to control for different income levels. My results show the advantage of controlling for pre-reform income non-parametrically, i.e., exclusively focusing on the MB group where the treated and control units have the almost identical pre-reform labor income distributions (Figure 1.2). This strategy makes results robust to inclusion of pre-reform income in the covariates. At the same time, however, my results also show that including log LI and its square is useful especially for the parallel pre-reform wage trends. Thus, I choose the bottom panel of Figure 1.3 as a main specification and present results based on this specification (i.e., Column 4 of Table 1.4) in the following analysis.

Before moving on to details of the main specification, I here demonstrate the advantage of focusing on the MB group again but I do so this time by comparing its wage dynamics with those of the other three groups, i.e., the BM, MT, and TM groups. Figure 1.4 shows wage dynamics of all the groups. Control units are weighted by entropy balancing as before, but in this figure only, I removed age square, household assets, and log LI

---

<sup>21</sup>As the difference is only in weights for the control units between the two panels, wage dynamics of the treated units are same in the two panels.

square from the covariates to balance. This is because for the other three groups, entropy balancing fails to find unique weights and thus one cannot balance covariates when these three variables are included. This failure typically happens when treated units are too different from control units.<sup>22</sup> As Figure 1.2 clearly shows, the distributions and supports of log LI differ by a large amount between treated and control units for the BM, MT, and TM groups. This fact indicates that treated units might not be comparable to control units in these three groups even after entropy balancing. The left panel of Figure 1.4 does not include any LI control in the covariates to balance, while the right panel includes log LI. One can find the following clear pattern: treatment effects of the BM, MT, and TM groups are unstable and even flipped once controlling for pre-reform labor income. For example, the difference in wage dynamics between the treated and control units of the BM group almost disappears in the right panel. For the TM group, the treated units show higher wage growth in the left panel, while the control units do in the right panel. On the other hand, the MB group again shows robust wage dynamics. Finally, note that the wage dynamics of Figure 1.3 is different from those of Figure 1.4 for the MB group because the latter does not include age square, household assets, and log LI square in the covariates to balance, implying that results of the MB group are robust not only to pre-reform labor income controls but also to other characteristics. This exercise indicates that I can identify the treatment effect for the MB group but not necessary for the other three groups.

Let me now come back to the main specification of this paper, i.e., the bottom panel of Figure 1.3, which corresponds to Column 4 of Table 1.4. The bottom panel of Figure 1.3 is non-parametric graphical evidence combined with covariate balancing and shows a clear pattern of wage dynamics. First, the pre-reform trends are identical between the two units, which is, for the years 1985 and 1986, a direct consequence of balancing the pre-reform wage changes ( $\Delta \log w$ ).<sup>23</sup> The treated units of the MB group, who are pushed downward from the middle to bottom brackets, display higher wage growth after the reform. In addition, one can find accumulating effects as the difference between the two units spreads out over time.

I here clarify the meaning and interpretation of accumulating effects of taxes on wages by using Figure 1.5. The figure plots fractions of workers located in the middle bracket (the top panel) and the bottom bracket (the bottom panel).<sup>24</sup> Let us focus on the top panel. Both treated and control units are in the middle bracket in 1986 by construction.

<sup>22</sup>For example, if treated units are only males and control units are only females, one cannot construct weights to balance gender ratios.

<sup>23</sup>One can find a similar idea in a synthetic control method where researchers match pre-treatment outcomes in levels for sufficiently long pre-treatment periods. (Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010); Abadie et al. (2015))

<sup>24</sup>Due to data issues, I cannot pin down individual bracket locations in 1983.

Although the treated units are pushed downward to the bottom bracket by the reform, this bracket movement is only mechanical or counterfactual, i.e., they are by definition  $M^{86}(z_{i86})\tilde{B}^{87}(z_{i86})$  using the notation in Section 1.4. Notice that their behavior and income are fixed at the 1986 level  $z_{i86}$ . Their actual brackets in 1987 can be thus different from the bottom bracket because of behavioral changes. For this reason, the fraction of workers in the middle bracket in 1987 is not equal to zero for the treated units. This is also the case for the control units. That is, since they stay in the middle bracket only in the mechanical or counterfactual sense ( $M^{86}(z_{i86})\tilde{M}^{87}(z_{i86})$ ), their actual brackets in 1987 are not necessary the middle bracket.

However, the two panels of Figure 1.5 show that, after the 1987 reform, the treated units are more likely to be in the bottom bracket on average while the control units are more likely to be in the middle bracket. In addition, their bracket locations are almost identical between the two units before the reform. This observation gives the following interpretation about accumulating effects: before the reform the two units face same tax rates and thus have same incentives for, e.g., on-the-job human capital accumulation, which contributes to the parallel pre-reform wage trends. They respond differently to taxes during 1987 due to the mechanical changes in brackets created by the reform, leading to the small initial difference in wages found in Figure 1.3. Regarding the actual bracket locations at the end of 1987, Figure 1.5 shows that the fraction of workers in the bottom bracket is still significantly higher for the treated units, which generates larger incentives for on-the-job human capital accumulation once again and leads to the widening gap of wages in 1988. This process continues over the years, and hence taxes create the accumulating effects on wages.

### 1.5.2 Threats to identification

Before moving on to regression analysis, it is worthwhile to consider potential threats to identification. As far as I am aware, strategic behavior to control income and composition changes in employed workers over time will be the main issues to be examined.

**Strategic behavior.** Since the identification strategy exploits bracket movements, one concern is strategic behavior that will lead to bunching around bracket cutoffs as found by Chetty et al. (2011b) and le Maire and Schjerning (2013) in Denmark. When close to a bracket cutoff, workers may try to control their income strategically to avoid crossing the bracket by, e.g., adjusting labor supply or shifting income. If this manipulation behavior is dominant, it will be misleading to interpret the accumulating effects of taxes on wages as a phenomenon resulting from real behavioral responses such as on-the-job human capital accumulation.



To look at this concern, I start from the popular observation that a degree of strategic income-control should be revealed by bunching around the bracket cutoff.<sup>25</sup> Figure 1.6 plots fractions of workers by income for the middle bracket over the three years after the reform. Income is measured in difference from the bracket cutoff and grouped into 1,000-DKK bins.<sup>26</sup> In 1987, for example, income for the middle bracket is  $PI + [CI > 0]$  (i.e., the tax base), and the cutoff for the middle bracket is 130,000 DKK as listed in Table 1.2. The figure confirms that both treated and control units display smooth densities and no spikes around the bracket cutoff. Given that the samples are male wage-earners, this finding is consistent with Chetty et al. (2011b) and le Maire and Schjerning (2013), who find evident bunching formed by female or self-employed workers. I hence conclude that strategic behavior to control income is not a serious threat to identification.

**Composition change.** Although Figure 1.3 presents the wage response to taxes, one has to bear in mind that, after the reform, workers can be non-employed and also can drop from the samples due to attrition caused by, e.g., death or emigration. Obviously, we cannot observe wages for non-employed or missing workers. Note that workers are employed in all of the pre-reform years (1983 - 1986) by sample selection. Figure 1.7 plots an attrition-adjusted employment rate in a year  $t$ , which is computed as the number of employed workers in a year  $t$  divided by the number of workers in 1986. This employment rate adjusts attrition because it is relative to workers observed in 1986, who are not subject to attrition and thus constitute the population of the samples. The figure shows that the employment rates of the treated and control units are almost same over time even after taking attrition into consideration.

Although the quantity of employed (i.e., the employment rate) is found to be same over time between the two units, there is still a concern that the quality of employed (i.e., the composition) changes disproportionately. Covariate balancing makes the two units similar in terms of the pre-reform 1986 covariates. However, if the reform affects incentives to become non-employed (e.g., self-employed) disproportionately for the two units, then the composition of employed workers will change through selection after the reform. This composition changes over time make it problematic to compare the two units because the parallel trend assumption in the DID design will be violated. Table 1.5 lists the pre-reform 1986 covariates of those who are employed in 1993, which is informative to see who remains employed after the reform in terms of the pre-reform covariates. One can confirm that the compositions of the two units are still balanced seven years after the reform. To put it more precisely, although the compositions of the two units do change over time

<sup>25</sup>The literature exploits bunching to identify an ETI. See Kleven (2016) for a survey.

<sup>26</sup>The bins are  $[-20000, -19000)$ , ...,  $[-1000, 0)$ ,  $(0, 1000]$ , ...,  $(19000, 20000]$ . If income is exactly equal to the bracket cutoff (i.e., the difference is zero), its observation is included in  $[-1000, 0)$ .

as is clear from comparing Table 1.4 with 1.5, they change in the same way between the two units. Then, wage dynamics resulting from the composition changes will be washed away by differencing out in the DID framework, implying that the composition changes in employed workers are not a threat to identification. In addition, the regression analysis of Section 1.5.3 controls for the pre-reform covariates in a parametric way.

### 1.5.3 Regression analysis

**Specification.** To further look into accumulating effects, I move on to regression analysis and adopt the following DID specification for an individual  $i$  and a year  $t = 86, \dots, 93$ :

$$\begin{aligned} \log w_{it} = & \alpha_0 + \alpha_1 1\{i \in T\} + \alpha_{87} 1\{t = 87\} + \dots + \alpha_{93} 1\{t = 93\} \\ & + \beta_{87}^w 1\{i \in T\} 1\{t = 87\} + \dots + \beta_{93}^w 1\{i \in T\} 1\{t = 93\} + X'_{i86} \gamma + u_{it}, \end{aligned} \quad (1.2)$$

where  $T$  is a treatment group indicator and  $u_{it}$  is an error term.  $t = 86$  is a reference year, and the parameters of interest are  $\beta_{87}^w$ , ..., and  $\beta_{93}^w$ , i.e., the treatment effects one year, ..., and seven years after the reform. Standard errors are clustered by individual (Bertrand, Duflo, and Mullainathan (2004)).

The control units are weighted by entropy balancing.  $X_{i86}$  is a vector of the pre-reform covariates and exactly same as the set of variables included in covariate balancing, i.e., the variables listed in Table 1.4. This control adjusts the tiny difference in the compositions between the two units after the reform, which we detected in Table 1.5. Note that controlling for post-reform covariates is “bad control” because they are in general endogenous to the reform and therefrom cause selection biases (Angrist and Pischke (2009)).

**Intertemporal Elasticity.** To put estimation results into perspective, it is necessary to transform  $\beta_{86}^w$ , ..., and  $\beta_{93}^w$  into elasticities of wages with respect to tax rates. Following the literature, I consider a net-of-marginal tax rate  $1 - \tau^t$ , where  $\tau^t$  is a marginal tax rate on labor income under a year- $t$  tax system. In order to incorporate the details of the Danish tax system, one should focus on effective (rather than statutory) marginal tax rates on labor income. Since effective tax rates are not available in the data set, I use the tax simulator and compute them by the following formula:

$$\tau^t(\text{LI}_{it}, y_{it}) = \frac{\text{Tax}^t(\text{LI}_{it} + 100, y_{it}) - \text{Tax}^t(\text{LI}_{it}, y_{it})}{100},$$

where  $\text{Tax}^t(\text{LI}_{it}, y_{it})$  is a tax liability under a year- $t$  tax system.  $\text{LI}_{it}$  is labor income, and  $y_{it}$  includes the other income components and individual characteristics such as marital status, i.e.,  $y_{it} = \{\text{OPI}_{i86}, \text{CI}_{i86}, \text{D}_{i86}, x_{i86}\}$  using the notation in Section 1.4.

As the treated and control units are defined by the mechanical changes in brackets between 1986 and 1987, mechanical changes in net-of-marginal tax rates are analogously defined by

$$\log(1 - \tilde{\tau}^{87}(\text{LI}_{i86}, y_{i86})) - \log(1 - \tau^{86}(\text{LI}_{i86}, y_{i86})),$$

where  $\tilde{\tau}^{87}(\text{LI}_{i86}, y_{i86})$  is a counterfactual marginal tax rate under the inflation-adjusted 1987 tax system. Note that individual behavior and income are fixed at the 1986 level  $(\text{LI}_{i86}, y_{i86})$ . Then, by running the following first-stage regression, I estimate differential mechanical changes in net-of-marginal tax rates between the treated and control units:

$$\log(1 - \tilde{\tau}^{87}(\text{LI}_{i86}, y_{i86})) - \log(1 - \tau^{86}(\text{LI}_{i86}, y_{i86})) = \delta_0 + \delta_1 1\{i \in T\} + u_i. \quad (1.3)$$

Given that  $\delta_0$  is an average for the control units and that  $\delta_0 + \delta_1$  is an average for the treated units,  $\delta_1$  captures the differential changes and thus is a proper measure for variation in tax rates.

Finally, one can compute intertemporal elasticities of wages with respect to net-of-marginal tax rates by dividing estimates of  $\beta_{87}^w$ , ..., and  $\beta_{93}^w$  by an estimate of  $\delta_1$  respectively, i.e.,  $\hat{\epsilon}_t^w := \hat{\beta}_t^w / \hat{\delta}_1$  for  $t = 87, \dots, 93$ . Standard errors are adjusted by the delta method treating  $\hat{\delta}_1$  as a constant variable. The elasticity  $\hat{\epsilon}_t^w$  is intertemporal in the sense that it measures the response in wages several years after the reform while the variation in tax rates takes place only at the reform. As it turns out in Section 1.6, this elasticity concept connects the reduced-form parameters to a welfare formula.

**Placebo.** To check pre-reform wage trends, I also conduct a placebo test using the pre-reform period ( $t = 83, \dots, 86$ ) as follows:

$$\begin{aligned} \log w_{it} = & \alpha_0 + \alpha_1 1\{i \in T\} + \alpha_{83} 1\{t = 83\} + \dots + \alpha_{85} 1\{t = 85\} \\ & + \beta_{83}^w 1\{i \in T\} 1\{t = 83\} + \dots + \beta_{85}^w 1\{i \in T\} 1\{t = 85\} + u_{it}. \end{aligned} \quad (1.4)$$

Same as in the main specification,  $t = 86$  is a reference year. The placebo treatment effects  $(\hat{\beta}_{83}^w, \dots, \hat{\beta}_{85}^w)$  must not be statistically different from zero to be confident that the wage trends are identical between the treated and control units in the absence of the treatment.<sup>27</sup> I also transform the DID estimates into wage elasticities by  $\hat{\epsilon}_t^w := \hat{\beta}_t^w / \hat{\delta}_1$  for

---

<sup>27</sup> Although I use the words “reform” and “treatment” interchangeably throughout the paper, one needs to be careful with the distinction here. Notice that both of the treated and control are affected by the reform. In other words, the treatment given to the treated units is not the reform but the mechanical movement from the middle to bottom brackets. Thus, the parallel trend assumption in this DID design is that the treated units have the same wage trends as the control units when the former hypothetically stays in the middle bracket in the mechanical sense.

$t = 83, \dots, 85$ .

Notice that, as opposed to the main specification given by Equation (1.2), I do not include the pre-reform covariates  $X_{i86}$  in Equation (1.4). Covariate balancing makes the treated and control units similar in terms of the pre-reform covariates. Further, workers are employed in all of the pre-reform years by sample selection. These points mean that there is no composition changes, and thus the two units are directly comparable before the reform. For this reason, one does not need to include the pre-reform covariates  $X_{i86}$  in Equation (1.4).

**Result.** Figure 1.8 plots point estimates of the wage elasticities with their 95% confidence intervals based on the main specification by Equation (1.2) and also based on the placebo test by Equation (1.4). One can see that the regression results are in line with graphical evidence given by Figure 1.3.

First of all, none of the placebo estimates ( $\hat{\epsilon}_{83}^w, \dots, \hat{\epsilon}_{85}^w$ ) are statistically different from zero. Given that, to deal with mean reversion, I include the wage changes between 1985 and 1986 ( $\Delta \log w = \log w_{i86} - \log w_{i85}$ ) in the covariates to balance (Table 1.4),  $\hat{\epsilon}_{85}^w$  is constrained to zero almost by construction. However, the fact that the other non-constrained placebo estimates ( $\hat{\epsilon}_{83}^w$  and  $\hat{\epsilon}_{84}^w$ ) are not statistically different from zero gives me some confidence that wage trends would be parallel in the absence of the treatment.

Next, the treatment effects ( $\hat{\epsilon}_{87}^w, \dots, \hat{\epsilon}_{93}^w$ ) are all statistically different from zero. That is, workers respond to taxes through wages. In addition, although the confidence intervals overlap, the point estimates tend to get large over time, which indicates accumulating effects. More specifically, the elasticity is around 0.05 one year after the reform but reaches around 0.1 five years after the reform. If distortion by taxes had only a static impact, these estimates would be same over time.

In relation to the literature, Blomquist and Selin (2010) also estimate the elasticity of hourly wages with respect to net-of-tax rates using Swedish survey data and register data. They consider a static specification, rather than the dynamic one considered here, while exploiting variation in tax rates over ten years. Their static elasticity estimates are in the range of 0.14 - 0.16 for males and thus of the same magnitude as my intertemporal elasticity estimates. One of the differences between Blomquist and Selin (2010) and the current paper is an identification strategy. In a regression framework widely adopted in the ETI literature, they attribute wage dynamics over ten years to time series variation in tax rates over ten years. On the other hand, my identification strategy exploits the specific tax reform and explicitly compares the treated to control units in the DID framework. Thanks to this strategy, I can follow the same workers over time and thus estimate the accumulating effects of taxes on wages, which are novel in the literature.

**Exclusion restriction.** Having presented the main result regarding taxes on wages, I come back to and address validity of an exclusion restriction, i.e., the identification assumption given by Equation (1.1). To test whether or not the other components of pre-reform income  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  affect wage dynamics conditional on  $LI_{i86}$  and demographics, I consider the following regression by treatment status for a year  $t = 86, \dots, 93$ :

$$\log(w_{it}) = \alpha_0 + \alpha_{87}1\{t = 87\} + \dots + \alpha_{93}1\{t = 93\} + X'_{i86}\gamma + u_{it}. \quad (1.5)$$

Under this specification, which is very similar to Equation (1.2), wage dynamics are captured by  $\alpha_{87}$ , ..., and  $\alpha_{93}$ . I then run the following three specifications with different controls for pre-reform income:

1.  $X_{i86}$  includes all the variables listed in Table 1.4 (the baseline control).
2.  $X_{i86}$  includes the baseline control and  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$ .<sup>28</sup>
3.  $X_{i86}$  excludes  $LI_{i86}$  controls (i.e, log LI and its square) from the baseline control.

If  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  do not affect wage dynamics, Specifications 1 and 2 must give statistically same estimates  $(\hat{\alpha}_{87}, \dots, \hat{\alpha}_{93})$ . I also test the importance of controlling for  $LI_{i86}$  by comparing estimates of Specification 1 to those of Specification 3.

Note that if one conducts the same exercise using Equation (1.2) rather than Equation (1.5), estimates between Specifications 1 and 2 will be different for the two indistinguishable reasons: First,  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  affect wage dynamics. Second, identification is destroyed because there is no variation in treatment assignment. To isolate the first reason from the second, I estimate Equation (1.5) separately for the treated and control units. Finally, since this exercise does not compare the two units, the control units are raw and not weighted by entropy balancing.

Figure 1.9 plots point estimates of the wage-dynamics coefficients  $(\hat{\alpha}_{87}, \dots, \hat{\alpha}_{93})$  with their 95% confidence intervals based on Equation (1.5). Same patterns appear for both treated and control units. One can find that the point estimates and confidence intervals are almost identical between Specification 1 (blue in the figure) and Specification 2 (red), which implies that exclusion restriction is verified. On the other hand, Specification 3 (green) gives the significantly different estimates from the other two specifications, meaning that labor income does affect wage dynamics. In conclusion, the figure justifies my identification strategy, which controls for labor income but does not control for the other income components for variation in treatment assignment.

---

<sup>28</sup>Since  $\{OPI_{i86}, CI_{i86}, D_{i86}\}$  can be zero or negative, I use them in levels not in logs.

### 1.5.4 On-the-job training

What is the underlying channel for the response to taxes though wages? There will be multiple answers, but this paper investigates a hypothesis that on-the-job human capital accumulation is one channel. Although workers can accumulate human capital in various ways, I exclusively look at on-the-job training and learning-by-doing because proper outcome variables are available for them. This section presents results about on-the-job training while Section 1.5.5 focuses on learning-by-doing.

For the measure of on-the-job training, I use information on whether or not a worker attends training courses co-sponsored by the government. In 1986, for example, 5.0 percent of the treated units and 7.6 percent of the (unweighted) control units attend courses. The courses consist of language classes, vocational training, and college-level higher education among others. Simonsen and Skipper (2008) and Sørensen and Vejlin (2014) point out that a large fraction of training is indeed co-sponsored by the government in Denmark.

Given the binary dependent variable, I estimate the following linear probability model in the DID framework:

$$\begin{aligned} 1\{OJT_{it}\} = & \alpha_0 + \alpha_1 1\{i \in T\} + \alpha_{87} 1\{t = 87\} + \dots + \alpha_{93} 1\{t = 93\} \\ & + \beta_{87}^{OJT} 1\{i \in T\} 1\{t = 87\} + \dots + \beta_{93}^{OJT} 1\{i \in T\} 1\{t = 93\} + X'_{i86} \gamma + u_{it}, \end{aligned} \quad (1.6)$$

where  $1\{OJT_{it}\}$  equals to one if an individual  $i$  attends training courses co-sponsored by the government in a year  $t$  and to zero otherwise.<sup>29</sup> The estimation procedure and specification of controls are same as the DID wage regression of Equation (1.2) except for the outcome variable. I also conduct a placebo test and transform all the DID estimates into OJT elasticities by  $\hat{e}_t^{OJT} := \hat{\beta}_t^{OJT} / \hat{\delta}_1$ .

Figure 1.10 plots point estimates of the OJT elasticities with their 95% confidence intervals. First, all the placebo estimates are not statistically different from zero. Next, all the treatment effects are positive but not statistically different from zero. Some placebo estimates almost entirely overlap with some treatment effects. Hence, there seems no significant effects of taxes on participation in training courses on the job. This is probably because a decision on taking courses will be made not solely by a worker but jointly with his employer. If he cannot freely choose whether to take courses, income taxes will not have impacts on the incidence of on-the-job training.

---

<sup>29</sup>Lechner (2010) shows that a non-linear model (e.g., a probit model) leads to an inconsistent estimator in a DID framework. This is intuitively because, as opposed to a linear model, the differences are not washed away by subtraction.

### 1.5.5 Learning-by-doing

Let us next move on to learning-by-doing as another channel of the wage response to taxes. Since it is difficult to directly observe learning-by-doing in the data, I look at a response in hours worked. The idea is that if an employee works longer, he is more likely to acquire skills and knowledge, which should lead to higher wages. As I explain in Section 1.3, the data set includes daily hours worked from 1985.

Using hour worked as an outcome variable, I adopt the same DID specification as Equation (1.2) with a different pre-reform covariate vector  $Y_{i86}$ :

$$\begin{aligned} \log h_{it} = & \alpha_0 + \alpha_1 1\{i \in T\} + \alpha_{87} 1\{t = 87\} + \dots + \alpha_{93} 1\{t = 93\} \\ & + \beta_{87}^h 1\{i \in T\} 1\{t = 87\} + \dots + \beta_{93}^h 1\{i \in T\} 1\{t = 93\} + Y_{i86}' \gamma + u_{it}. \end{aligned} \quad (1.7)$$

Given that the relationship considered here is taxes on hours worked, I choose the pre-reform covariate vector  $Y_{i86}$  by following the labor supply literature, which emphasizes controlling for wages and assets in an intertemporal setting (Keane (2011)).  $Y_{i86}$  thus includes demographics in 1986 (same demographics included in  $X_{i86}$ ), hour changes between 1985 and 1986 ( $\Delta \log h = \log h_{i86} - \log h_{i85}$ ), and  $\log w$  in 1986.<sup>30</sup> Now that I introduce a new set of covariates, I redo covariate balancing and present results in Table 1.6, which again shows that entropy balancing works well to achieve high balance. In this section only, entropy balancing is based on not  $X_{i86}$  but  $Y_{i86}$ .

Same as before, I conduct a placebo test and transform all the DID estimates into hour elasticities by  $\hat{\epsilon}_t^h := \hat{\beta}_t^h / \hat{\delta}_1$ . Figure 1.11 plots point estimates of the hour elasticities with their 95% confidence intervals. The result is consistent with findings of the labor supply literature: that is, the intertemporal elasticities of males are small and sometimes insignificant (Meghir and Phillips (2010)). However, interestingly, Figure 1.11 shows an increasing pattern of the hour elasticities as we found it for the wage elasticities in Figure 1.8. Starting from around zero, the hour elasticity reaches around 0.15 seven years after the reform.

What is an interpretation of the increasing patterns of the wage and hour elasticities? One can think of the following feedback effect between wages and hours through learning-by-doing as one of the interpretations: initially, a worker responds to tax changes by, e.g., undertaking more on-the-job training or putting more unobservable work effort, leading to higher wages. This appears as the wage elasticity one year after the reform in Figure 1.8. Given higher wages, he has an incentive to work longer, which appears as the increasing hour elasticity after the reform in Figure 1.11. Then, through learning-by-doing, his

<sup>30</sup>The motivation of controlling for  $\Delta \log h$  is again to deal with mean reversion.

wages become even higher (Figure 1.8), which in turn creates an incentive to work even longer (Figure 1.11). Therefore, the increasing patterns of the wage and hour elasticities are consistent to each other if one has learning-by-doing in mind as a potential channel. This interpretation implies the wage response is not independent of the hour response; rather, they are interdependent through learning-by-doing. To further look into this joint relationship, it will be useful to see, for example, whether taxes have an impact on a joint probability of earning higher wages *and* working longer hours.

In relation to the literature, Keane (2015, 2016) structurally estimates life-cycle labor supply models with human capital accumulation via learning-by-doing. He also investigates an effect of taxes on male hours worked as well as wages and finds increasing wage and hour responses over time by simulation exercises. The underlying mechanism is same as my feedback effect between wages and hours explained in the previous paragraph. The current paper is different from Keane (2015, 2016) in terms of empirical approaches. While I exploit the tax reform as a natural experiment to estimate the reduced-form parameters, he constructs and estimates the models structurally. Although his structural estimates are much larger than my reduced-form estimates, these two estimates give the qualitatively same findings and come to the similar conclusions.<sup>31</sup> My reduced-form estimates are the first and direct empirical evidence in the literature that suggests that learning-by-doing is one of the underlying channels, and thus will be a basis or justification for exploring this channel using structural models as is done by Keane (2015, 2016). Hence, one can think of his papers and mine as complements to each other.

## 1.6 Welfare analysis

The literature of labor supply and an ETI typically considers a welfare consequence of income taxes under a static labor supply model with exogenous wages. However, the previous section finds the empirical evidence of accumulating effects of taxes on wages, which could change our understanding on welfare. Therefore, this section shifts an attention to welfare by constructing a simple model as a conceptual framework. Rather than specifying and estimating the fully parametric model, I use the model as a guide that links the elasticity estimates to welfare implications.

---

<sup>31</sup>One potential reason for his large estimates is that his samples are from NLSY79 and thus young American workers. It is widely reported that young workers have higher wage growth, which implies that they may respond more to tax changes through wages.



### 1.6.1 Model

To compare a welfare impact of wage responses with that of labor supply responses, I extend a canonical labor supply model to include choice on human capital accumulation on the job in a dynamic setting. Although the models by Chetty (2009a) and Feldstein (1999) feature sheltering behavior (i.e., tax avoidance or evasion) in addition to real behavior (i.e., labor supply), I omit it here because there is almost no opportunity for sheltering left for the majority of employed workers in Denmark (Kleven (2014)).

The economy consists of a unit measure of workers. Time is discrete, finite, and in years ( $t = 87, \dots, 93, \dots, T$ ).  $t = 87$  corresponds to the first year after the reform, while  $t = 93$  corresponds to the final year of the empirical analysis. One can regard  $T$  as a deterministic retirement period. A homogenous worker has initial and exogenous human capital  $k_{86}$  at the beginning of 87. It is straightforward to relax the assumptions on homogeneous and single-dimension human capital, which gives the same welfare implication. He then chooses continuous  $h_{87}$  hours to work, namely labor supply along the intensive margin. Given the samples include only males strongly attached to the labor market, I do not model decisions on labor market participation, i.e., labor supply along the extensive margin. Let  $\psi(h_{87})$  be an increasing and convex cost function of labor supply. He also invests  $i_{87}$  amounts of money in on-the-job training with an increasing and convex cost function  $c(i_{87})$ . On-the-job training should be broadly interpreted and includes not only training courses co-sponsored by the government but also informal training at workplaces.

Human capital accumulation takes place through either learning-by-doing or on-the-job training as follows: during the year ( $t = 87$ ), human capital  $k_{86}$  is upgraded to  $k_{87}$  following a law of motion  $k_{87} = F(h_{87}, i_{87}, k_{86})$ , where  $F(\cdot, \cdot, \cdot)$  is a human capital production function with every argument increasing and concave. Hourly wages are given by  $w_{87} = w(k_{87})$ , which is increasing and concave in  $k_{87}$ . The model is partial equilibrium in the sense that firm behavior and thus wage determination are exogenous. By assuming a constant (flat) marginal tax rate  $\tau$ , his net income at the end of 87 is given by  $(1 - \tau)w_{87}h_{87}$ .

He repeats the action  $(h_t, i_t)$  over the period ( $t = 87, \dots, 93, \dots, T$ ). I assume a zero discount rate and quasi-linear utility, which implies that there is no incentive for private savings. Let  $V(k_{86})$  denote a worker's value at the beginning of 87 with his initial human

capital  $k_{86}$ . Then, his problem is written as

$$\begin{aligned} V(k_{86}) = & \max_{\{h_t, i_t\}_{t=87}^T} (1 - \tau)w_{87}h_{87} - \psi(h_{87}) - c(i_{87}) \\ & + \cdots + \left\{ (1 - \tau)w_{93}h_{93} - \psi(h_{93}) - c(i_{93}) \right\} \\ & + \cdots + \left\{ (1 - \tau)w_T h_T - \psi(h_T) - c(i_T) \right\}, \end{aligned}$$

subject to

$$\begin{aligned} w_t &= w(k_t) \\ k_t &= F(h_t, i_t, k_{t-1}). \end{aligned}$$

On the other hand, tax revenue from him over the period is given by

$$R(k_{86}) = \tau w_{87}h_{87} + \cdots + \tau w_{93}h_{93} + \cdots + \tau w_T h_T$$

As is standard in the literature, the government is assumed to distribute its revenue as a lump-sum transfer among taxpayers. Therefore, given the quasi-linear utility, the (money-metric) social welfare is defined as

$$W = V(k_{86}) + R(k_{86}).$$

The interest lies on marginal excess burden of taxation when the government permanently changes the tax rate at the beginning of 87, i.e.,  $\frac{dW}{d\tau}|_{87}$ . By exploiting the envelope theorem, this welfare measure is reduced to

$$\begin{aligned} \frac{dW}{d\tau}\Big|_{87} &= \frac{dV(k_{86})}{d\tau}\Big|_{87} + \frac{dR(k_{86})}{d\tau}\Big|_{87} \\ &= -w_{87}h_{87} - \cdots - w_T h_T \\ &\quad + \left( w_{87}h_{87} + \tau \frac{dw_{87}h_{87}}{d\tau}\Big|_{87} \right) + \cdots + \left( w_T h_T + \tau \frac{dw_T h_T}{d\tau}\Big|_{87} \right) \\ &= w_{87}h_{87}(\epsilon_{87}^w + \epsilon_{87}^h) + \cdots + w_{93}h_{93}(\epsilon_{93}^w + \epsilon_{93}^h) + \cdots + w_T h_T(\epsilon_T^w + \epsilon_T^h), \end{aligned} \quad (1.8)$$

where  $\epsilon_t^w := \frac{dw_t/w_t}{d\tau/\tau}\Big|_{87}$  and  $\epsilon_t^h := \frac{dh_t/h_t}{d\tau/\tau}\Big|_{87}$  for  $t = 87, \dots, 93, \dots, T$ . This welfare formula is a straightforward extension of famous Feldstein's one to the dynamic setting and conveys the same insight: a small subset of estimable parameters forms sufficient statistics for welfare analysis. In particular, the intertemporal wage and labor supply elasticities ( $\epsilon_t^w$  and  $\epsilon_t^h$ ) capture behavioral responses and play central roles in the formula. Thus, the

model laid out here works as a guide for welfare analysis and gives me a mapping from the elasticity estimates to welfare implications.<sup>32</sup>

### 1.6.2 Clarification

Since the welfare analysis of this section is closely related to the one of the ETI literature, it is worthwhile to clarify some points. First of all, by definition, labor income is given by  $LI_t = w_t h_t$ , which immediately implies that  $\epsilon_t^{LI} = \epsilon_t^w + \epsilon_t^h$ . Thus, in order to apply the welfare formula of Equation (1.8),  $\epsilon_t^{LI}$  is sufficient. For this reason, a large amount of the literature estimates an ETI or an elasticity of labor income. They typically find large ETIs for top-income or self-employed workers and consider tax avoidance or evasion as potential behavioral responses (Kleven (2014) and Saez et al. (2012b)). The policy implications are that the government should fix the tax system by, e.g., removing loopholes. On the other hand, ETIs are small for the majority of employed workers, which implies that they may respond to taxes through real behavior (such as labor supply) rather than tax avoidance or evasion.

As Slemrod (1992, 1995, 1996, 1998) points out, real behavior is least responsive to taxes compared to avoidance or evasion but the most crucial factor for economic performance and growth. In addition, the anatomy of behavioral responses helps to better understand optimal policies because normative implications are different depending on behavioral responses. However, just looking at the aggregate response  $\epsilon_t^{LI}$  does not tell us its underlying channel. In other words, one needs to decompose the aggregate response into its micro responses to gain further welfare implications. In this paper, the decomposition is into the wage and hour responses ( $\epsilon_t^w$  and  $\epsilon_t^h$ ).

Although wages increase through learning-by-doing or on-the-job training in the model, one does not have to distinguish each response. Rather, what matters is only the wage response. As I decompose the labor income response into the wage and labor supply responses, further decomposition of the wage response into its roots is an important future work.

---

<sup>32</sup>The welfare formula of Equation (1.8) is valid when a reform is small. For a large reform like the 87 one, Kleven (2018) shows that the elasticity remains the sufficient statistics under quasi-linear and iso-elastic utility.

### 1.6.3 Welfare implications

Figure 1.12 stacks point estimates of the wage and hour elasticities ( $\hat{\epsilon}_t^w$  and  $\hat{\epsilon}_t^h$ ) estimated in Section 1.5 for  $t = 87, \dots, 93$ .<sup>3334</sup> Two points are worth mentioning. First, the wage responses have larger impacts on welfare than the labor supply responses for most of the years. Second, as summation of the two responses increases over time, the accumulating effects are evident. Given that the standard tool for analyzing tax incidence or optimal taxation is a static labor supply model with exogenous wages, the current results indicate that researchers should incorporate wage responses and dynamic aspects into their frameworks.

One can also gain further implications regarding tax policies aimed at encouraging labor supply. One such example is Earned Income Tax Credit (EITC) being implemented in US. EITC has a flat pyramid structure with phase-in (negative), flat (zero), and phase-out (positive) marginal tax rates. Clearly, phase-out tax rates create disincentives for labor supply. According to Figure 1.12, however, this negative impacts on labor supply and welfare could be mitigated by other tax policies encouraging human capital accumulation on the job. One can think of expanding deductions for on-the-job training costs as an example. This is because such policies will bring high wages and thus create incentives for labor supply. Although this is a simple example illustrating the benefit of looking at the micro responses ( $\epsilon_t^w$  and  $\epsilon_t^h$ ), one will not be able to gain the same implications just by looking at the aggregate response  $\epsilon_t^{LI}$ .

Note that  $\epsilon_t^{LI} = \epsilon_t^w + \epsilon_t^h$ , and  $\epsilon_t^{LI}$  has been estimated by Chetty et al. (2011b) and Kleven and Schultz (2014a) with Danish administrative data. In particular, Kleven and Schultz (2014a) use the same tax reform as the current paper and estimate  $\epsilon_{89}^{LI}$  taking the three-year difference (86 - 89). Although their elasticity concept is static rather than intertemporal, their estimate still gives me a useful benchmark. Their static elasticity estimate  $\hat{\epsilon}_{89}^{LI}$  is around 0.2 (reported as “DD elasticity” in their Panels A and B of Figure 4), which is slightly larger than my intertemporal elasticity estimate  $\hat{\epsilon}_{89}^w + \hat{\epsilon}_{89}^h \approx 0.13$ . This is partly because their samples include self-employed workers and top-income earners, who have larger elasticities as confirmed in their Table 4.<sup>35</sup> Given that my micro elasticity estimates ( $\hat{\epsilon}_{89}^w + \hat{\epsilon}_{89}^h$ ) are largely consistent with the aggregate elasticity estimate  $\hat{\epsilon}_{89}^{LI}$  of

<sup>33</sup>Notice that the elasticity estimates in Section 1.5 are defined with respect to net-of-marginal tax rates while the elasticities in Equation (1.8) are defined with respect to marginal tax rates. However, one can easily transform the latter into the former by multiplying  $-\tau/(1 - \tau)$ , which ensures that this small difference in definitions does not matter.

<sup>34</sup>Since the point estimate of  $\hat{\epsilon}_{87}^h$  is negative, its absolute value is stacked in the figure, which does not change any messages of this exercise.

<sup>35</sup>Another reason is that labor income in the data is observed as an yearly total amount of earnings while hourly wages and daily hours worked are observed only for workers holding jobs on the 28th of November. Therefore,  $\epsilon_t^{LI} = \epsilon_t^w + \epsilon_t^h$  holds only approximately.

Kleven and Schultz (2014a), it is worthwhile to clarify the difference between these two papers here: They estimate the aggregate elasticity ( $\hat{\epsilon}_{89}^L$ ), while I decompose it into or estimate the micro elasticities ( $\hat{\epsilon}_{89}^w, \hat{\epsilon}_{89}^h$ ). By estimating the micro elasticities, I can build the bridge between the labor supply and ETI literature. In particular, my findings imply that the response in (taxable) labor income ( $\hat{\epsilon}_{89}^L$ ) is explained not by the labor supply response ( $\hat{\epsilon}_{89}^h$ ) alone but jointly by the labor supply and wage responses ( $\hat{\epsilon}_{89}^w, \hat{\epsilon}_{89}^h$ ), which also has important welfare implications as shown in Figure 1.12.

## 1.7 Conclusion

This paper presents the micro evidence on accumulating effects of taxes on wages by exploiting the administrative data and tax reform in Denmark. Here, I mention a few robustness checks that I am currently working on.

First, I am looking at a concern that the wage response to taxes is due to shifting income from labor income to capital income rather than due to real behavior. One can think of stock options as a way to get salary in the form of capital income. Note that the reform changed the tax base for the middle bracket (Table 1.2). Taxpayers in the middle bracket with negative capital income have a large incentive to shift income from labor income to capital income after the reform because capital income is not included in the tax base for the middle bracket as long as it is negative. This income shifting creates lower wage growth for the control units and thus an upward bias in treatment effects. Although I conjecture that stock options are not common among non-executive workers in 1980's, this will be a valid concern.

Second, I am looking at another concern that the hour response to taxes is through job changes as found by Altonji and Paxson (1992) and Blundell, Brewer, and Francesconi (2008). Note that, in many cases, job changes are associated with wage increases as documented by Jolivet, Postel-Vinay, and Robin (2006) and Topel and Ward (1992). Then, it is misleading to interpret the hour response as evidence of learning-by-doing because one cannot distinguish between learning-by-doing and job changes regarding potential channels of the wage response. However, at the same time, it will be interesting to find some evidence that the wage response is through job changes.

Table 1.1: Income concept in the Danish tax system

Income concept	Acronym	Main items included
Taxable income	TI	PI + CI – D
Personal income	PI	LI + OPI
Labor income	LI	Salary, wages, bonuses, fringe benefits
Other personal income	OPI	Transfers – pension contributions
Capital income	CI	Interest income – interest on debt
Deductions	D	Commuting, union fees, UI contributions

Notes: Capital income is negative for the majority of Danish taxpayers as a result of interest payments on debt such as mortgage and other loans.

Table 1.2: Danish tax system before and after the 1987 reform

Tax type	1986			1987		
	Base	Rate	Cutoff	Base	Rate	Cutoff
Regional taxes	TI	28.0	20,700	TI	29.0	21,200
National taxes						
Bottom bracket	TI	19.9	23,200	TI	22.0	27,100
Middle bracket	TI	14.4	113,400	PI + [CI>0]	6.0	130,000
Top bracket	TI	10.8	186,100	PI	12.0	200,000

Notes: All the monetary values are in Danish Krone (DKK). DKK 1 in 1986  $\approx$  USD 0.3 in 2016. Regional taxes include municipal, county, and church taxes. Church taxes are paid only by members of the church (Folkekirken). The regional tax rates are averages across municipalities. The bottom tax rate in 1986 includes social security contributions.

Table 1.3: Descriptive statistics in 1986

Demographics:		Taxable income:	
Age	38.8	Labor income	186,042
Married (%)	59.0	Other personal income	4,740
Number of children	0.73	Capital income	-27,996
Low education (%)	31.3	Deductions	10,680
Middle education (%)	52.7		
High education (%)	16.0	Tax brackets:	
Household assets	226,318	Bottom bracket (%)	25.6
		Middle bracket (%)	54.5
Number of obs.	886,924	Top bracket (%)	19.3

Notes: Selected samples are male workers employed in all of the pre-reform years (1983 - 1986). Mean values as of 1986 are listed, and DKK 1 in 1986  $\approx$  USD 0.3 in 2016. Since taxable income of some workers is less than the bottom bracket cutoff, the three tax-bracket percentages do not add up to 100.

Table 1.4: Pre-treatment covarites (MB group in 1986)

Column	1	2	3	4
	Treated	Control		
Variable		Raw	Balanced	Balanced
Age	46.2	36.9	46.2	46.2
Age (square)	2259	1489	2259	2259
Married (%)	94.7	42.3	94.7	94.7
Number of children	0.83	0.60	0.83	0.83
Low education (%)	45.0	32.2	45.0	45.0
Middle education (%)	50.0	57.6	50.0	50.0
High education (%)	5.0	10.2	5.0	5.0
Household assets	256,318	154,532	256,270	256,262
$\Delta \log w$	0.008	0.037	0.008	0.008
$\log \text{LI}$	11.95	12.03		11.95
$\log \text{LI (square)}$	142.9	144.6		142.9
Number of obs.	107,985	281,727		

Notes: Pre-treatment covarites of the MB group in 1986 are listed, where  $\Delta \log w = \log w_{i86} - \log w_{i85}$ . Columns 1 and 2 display raw mean values by treatment status. Columns 3 and 4 display mean values of the control units weighted by entropy balancing.

Table 1.5: Pre-reform 1986 covarites (employed in 1993)

Variable	Treated	Control
Age	41.4	41.3
Age (square)	1,804	1,794
Married (%)	93.8	94.3
Number of children	1.13	1.12
Low education (%)	39.1	39.3
Middle education (%)	54.8	54.4
High education (%)	6.1	6.3
Household assets	222,318	213,482
$\Delta \log w$	0.013	0.012
$\log \text{LI}$	11.97	11.98
$\log \text{LI (square)}$	143.4	143.6

Notes: Pre-reform 1986 covarites of those employed in 1993 are listed. Control units are weighted by entropy balancing.

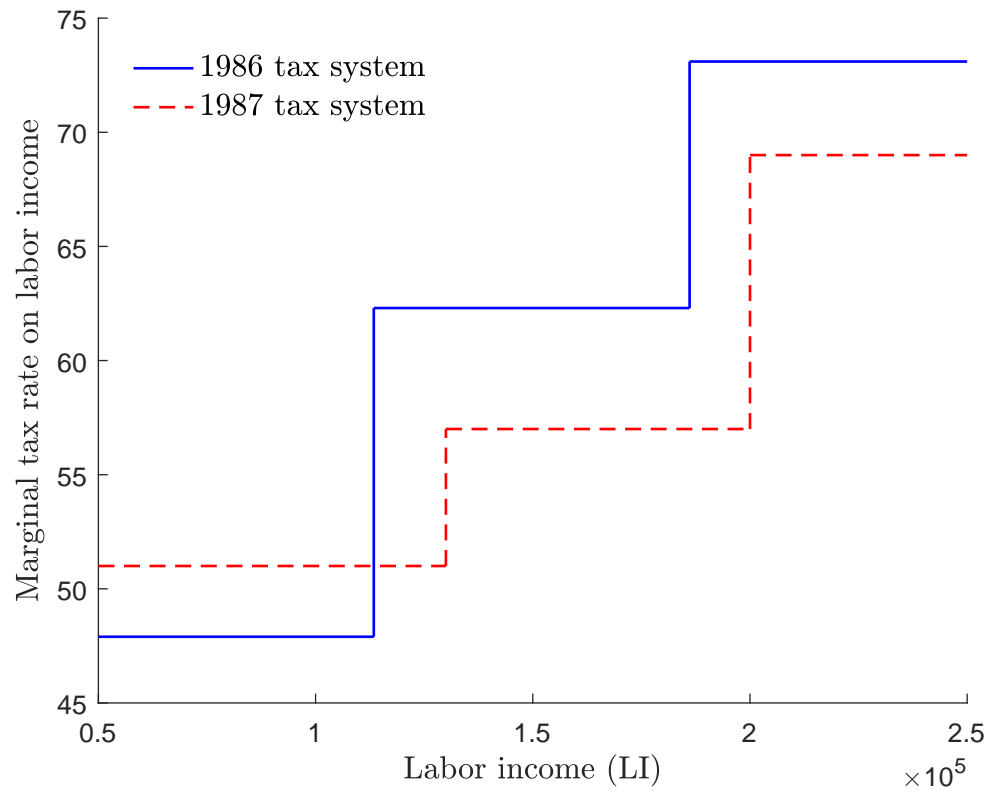
Table 1.6: Pre-treatment covarites (MB group in 1986)

Column	1	2	3
Variable	Treated	Control	
		Raw	Balanced
Age	46.2	36.9	46.2
Age (square)	2259	1489	2259
Married (%)	94.7	42.3	94.7
Number of children	0.83	0.60	0.83
Low education (%)	45.0	32.2	45.0
Middle education (%)	50.0	57.6	50.0
High education (%)	5.0	10.2	5.0
Household assets	256,318	154,532	256,268
$\Delta \log h$	0.008	0.017	0.008
$\log w$	4.52	4.57	4.52
Number of obs.	107,985	281,727	

Notes: Pre-treatment covarites of the MB group in 1986 are listed, where  $\Delta \log h = \log h_{i86} - \log h_{i85}$ . Columns 1 and 2 display raw mean values by treatment status. Columns 3 displays mean values of the control units weighted by entropy balancing.

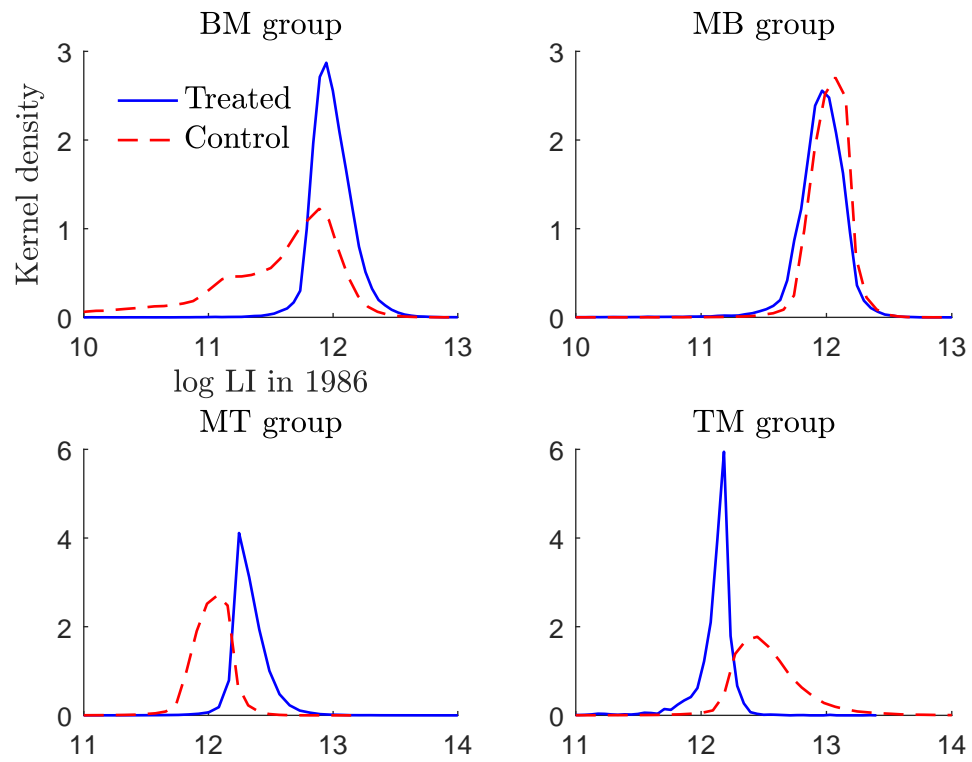


Figure 1.1: Overview of the 1987 tax reform



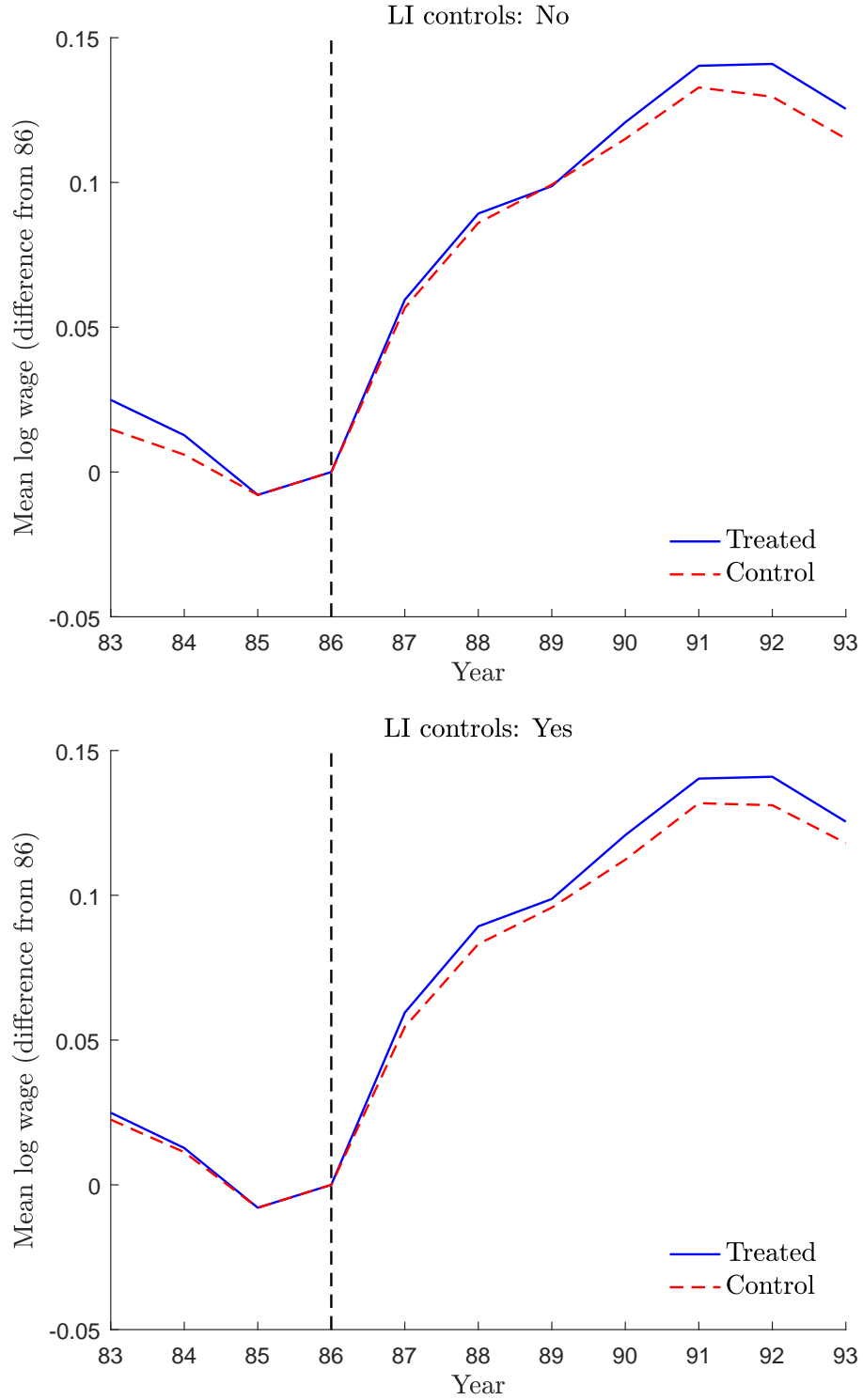
Notes: The figure plots marginal tax rates on labor income (LI) as a function of LI before and after the 1987 reform. The tax rates and cutoffs are from Table 1.2 while assuming other personal income (OPI), capital income (CI), and deductions (D) are all zero.

Figure 1.2: Correlation between treatment assignment and  $\log LI_{i86}$



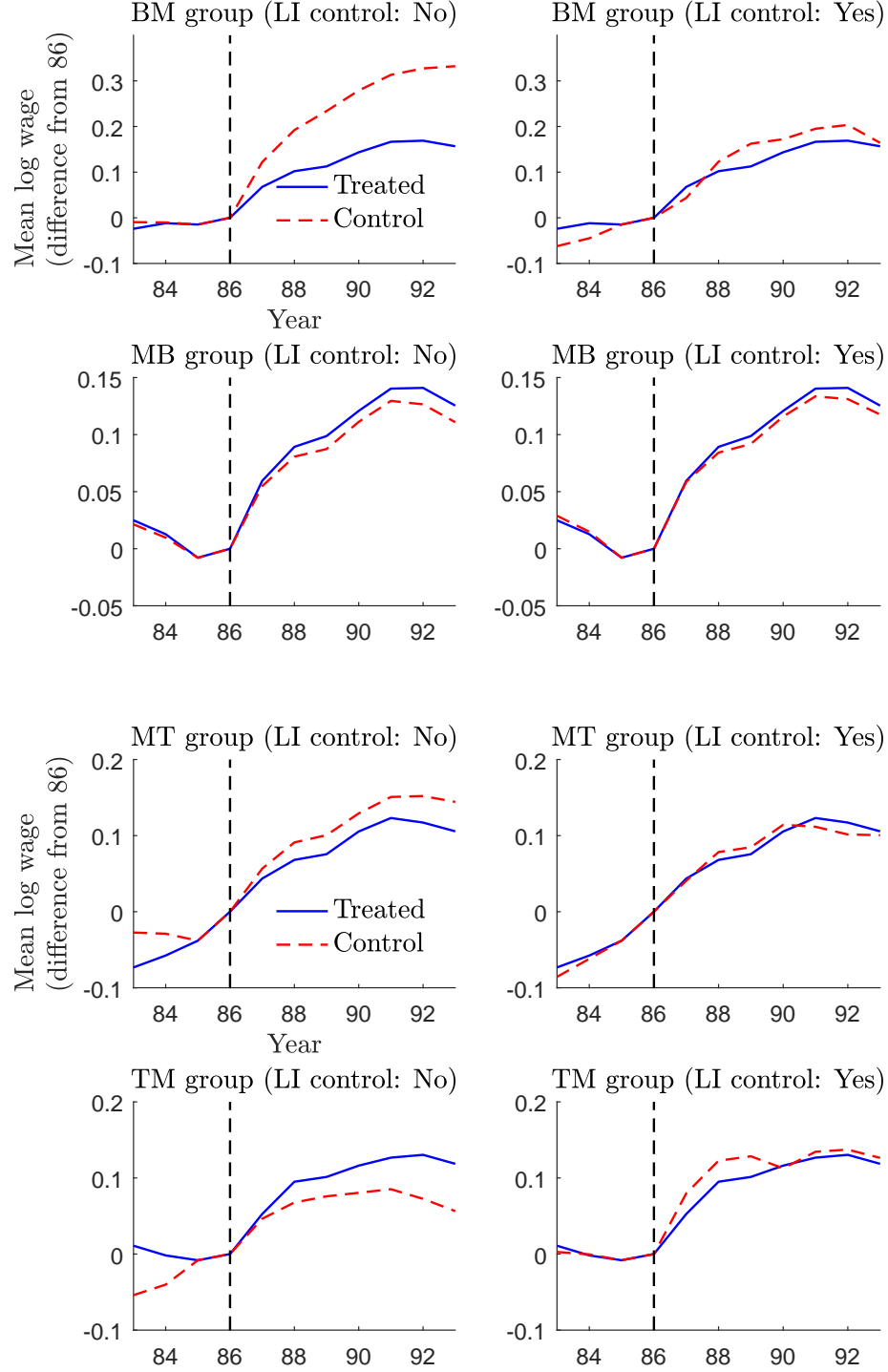
Notes: The figure plots kernel density estimates of  $\log LI_{i86}$  by treatment status for the BM, MB, MT, and TM groups. The estimation is based on `ksdensity` function in MATLAB with default settings.

Figure 1.3: Graphical evidence



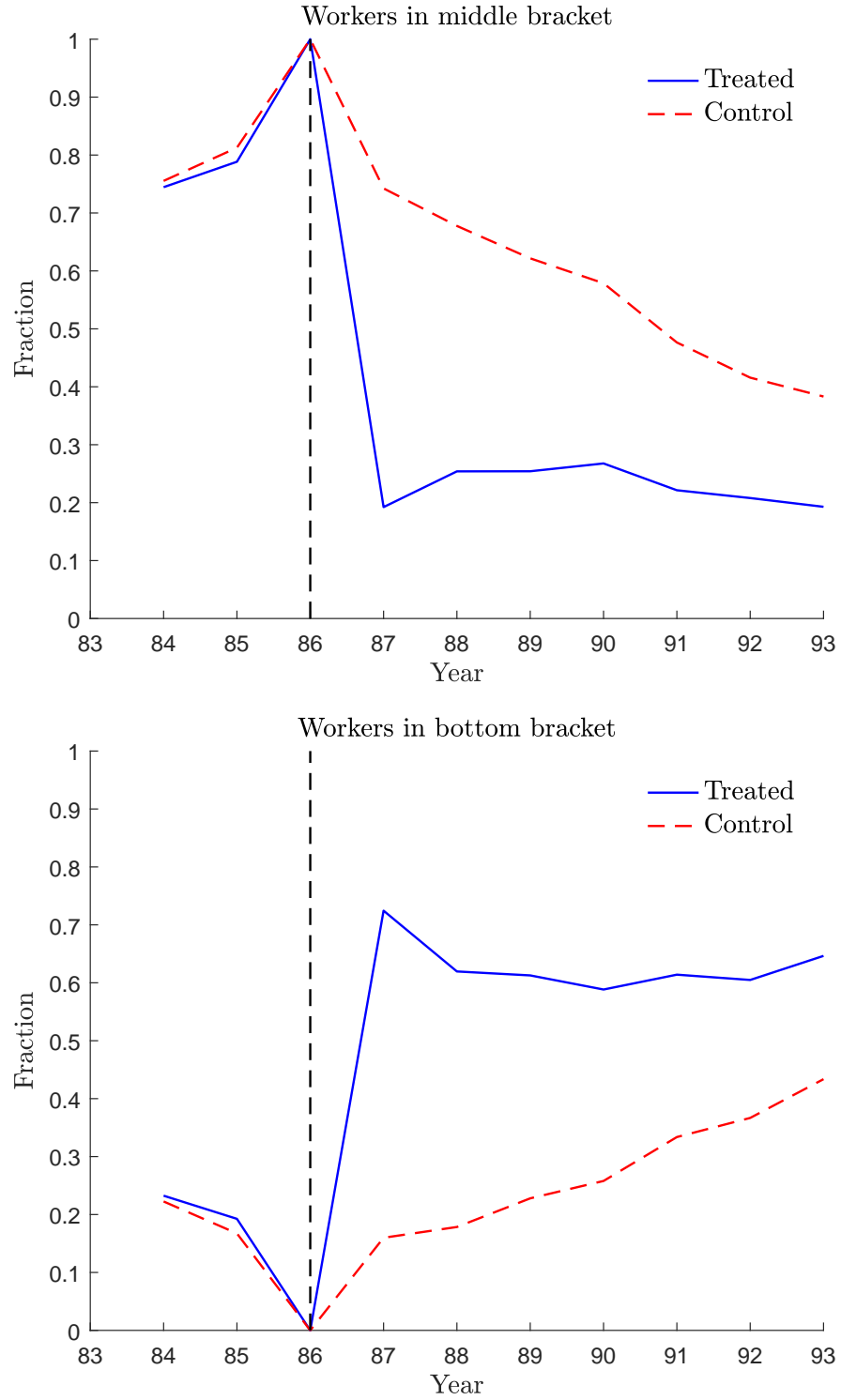
Notes: The figure plots  $\overline{\log w_{it}} - \overline{\log w_{i86}}$  for  $t = 83, \dots, 93$ . Control units are weighted by entropy balancing. The top panel does not include any LI controls in the covariates to balance (corresponding to Column 3 of Table 1.4) while the bottom panel includes log LI and its square (corresponding to Column 4 of Table 1.4).

Figure 1.4: Robustness of MB group



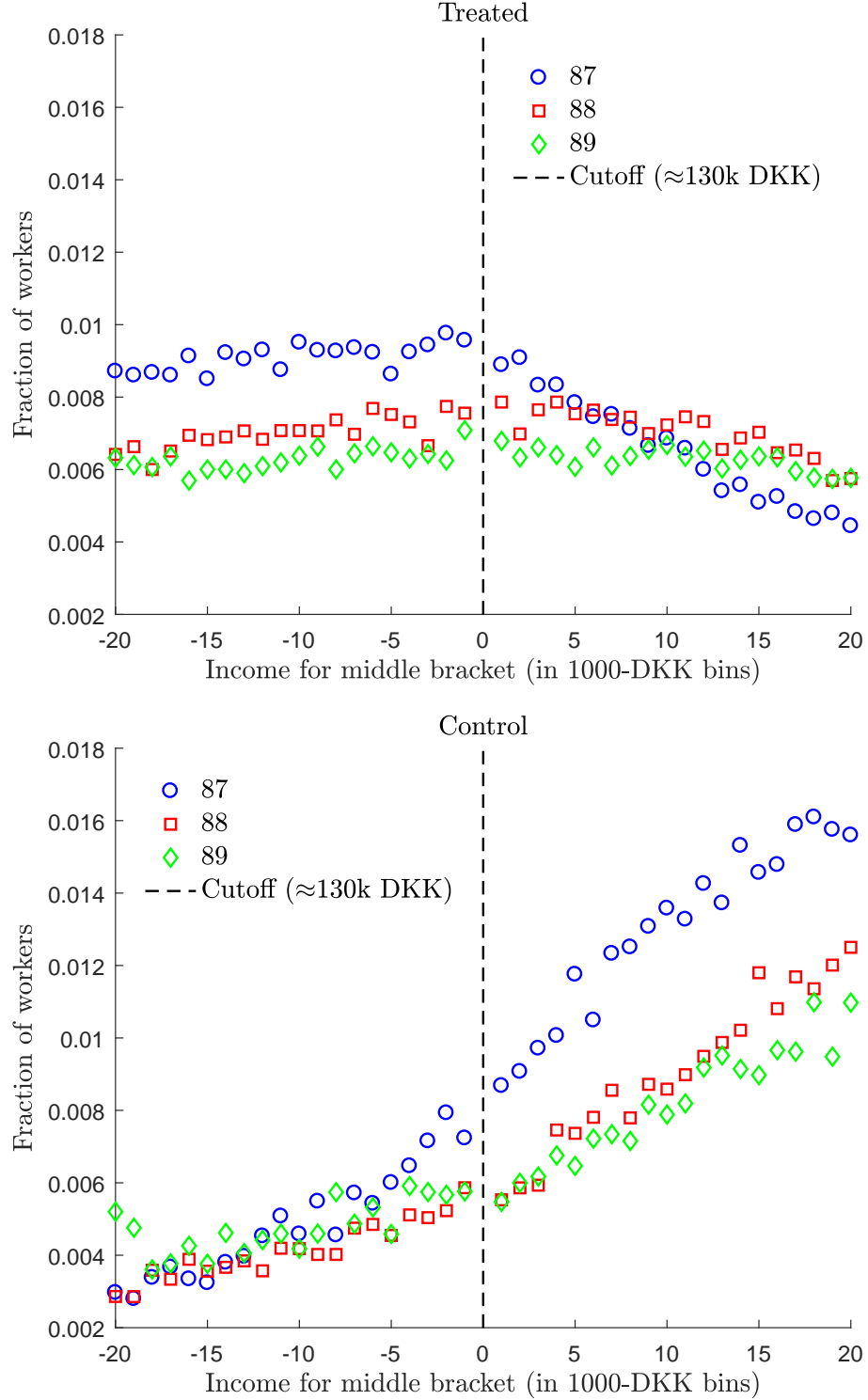
Notes: The figure plots  $\overline{\log w_{it}} - \overline{\log w_{i86}}$  for  $t = 83, \dots, 93$ . Control units are weighted by entropy balancing. The left panel does not include any LI control in the covariates to balance while the right panel includes log LI.

Figure 1.5: Actual tax brackets of MB group



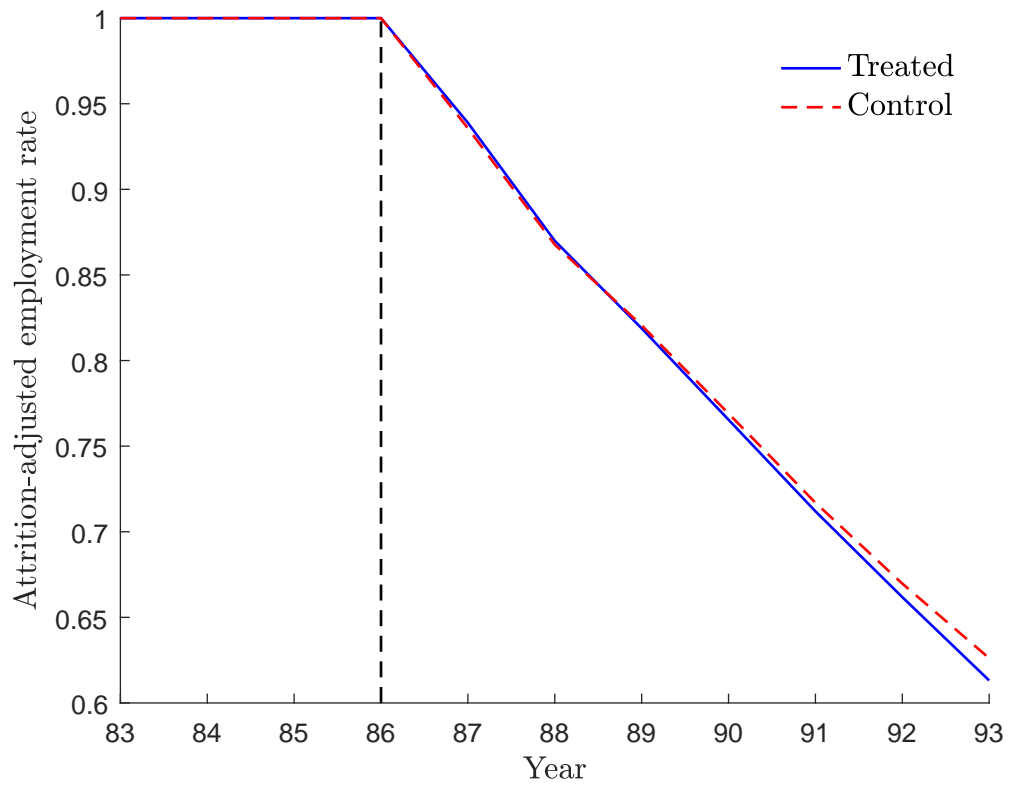
Notes: The figure plots fractions of workers located in the middle bracket (the top panel) and the bottom bracket (the bottom panel). Control units are weighted by entropy balancing.

Figure 1.6: Density around the middle bracket cutoff



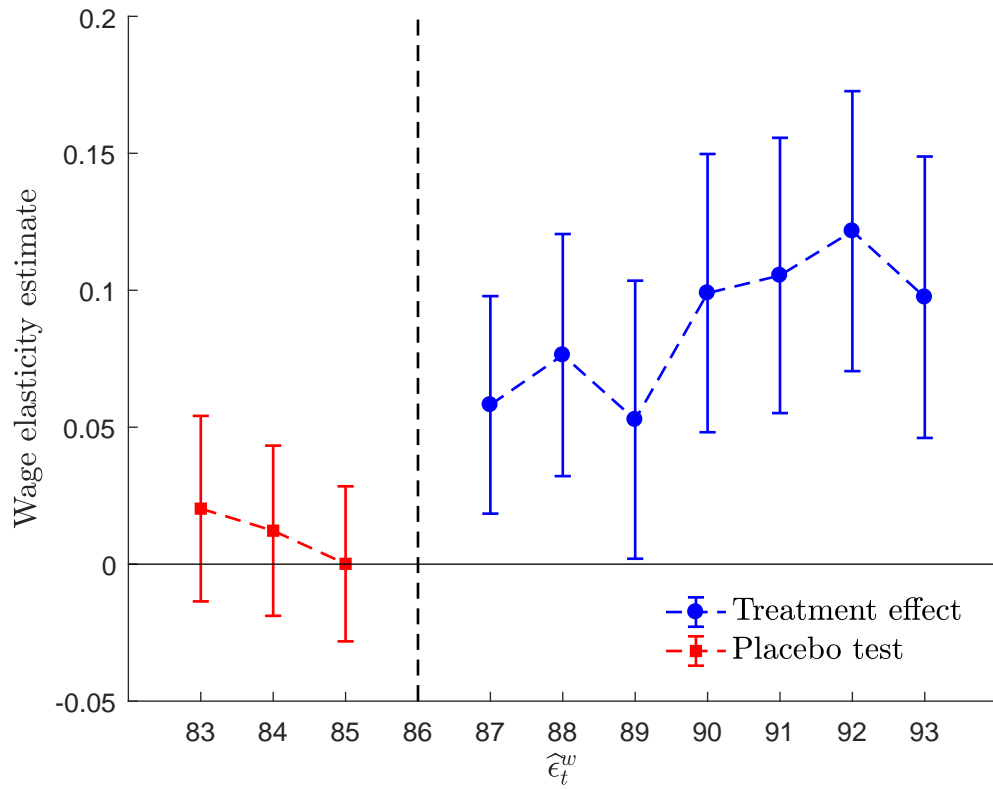
Notes: The figure plots fractions of workers by income for the middle bracket. Income is measured in difference from the bracket cutoff and grouped into 1,000-DKK bins. In 1987, for example, income for the middle bracket is  $PI + [CI > 0]$  (i.e., the tax base), and the cutoff for the middle bracket is 130,000 DKK. Control units are weighted by entropy balancing.

Figure 1.7: Composition change in employed workers



Notes: The figure plots an attrition-adjusted employment rate in a year  $t$ , which is computed as the number of employed workers in a year  $t$  divided by the number of workers in 1986. Control units are weighted by entropy balancing.

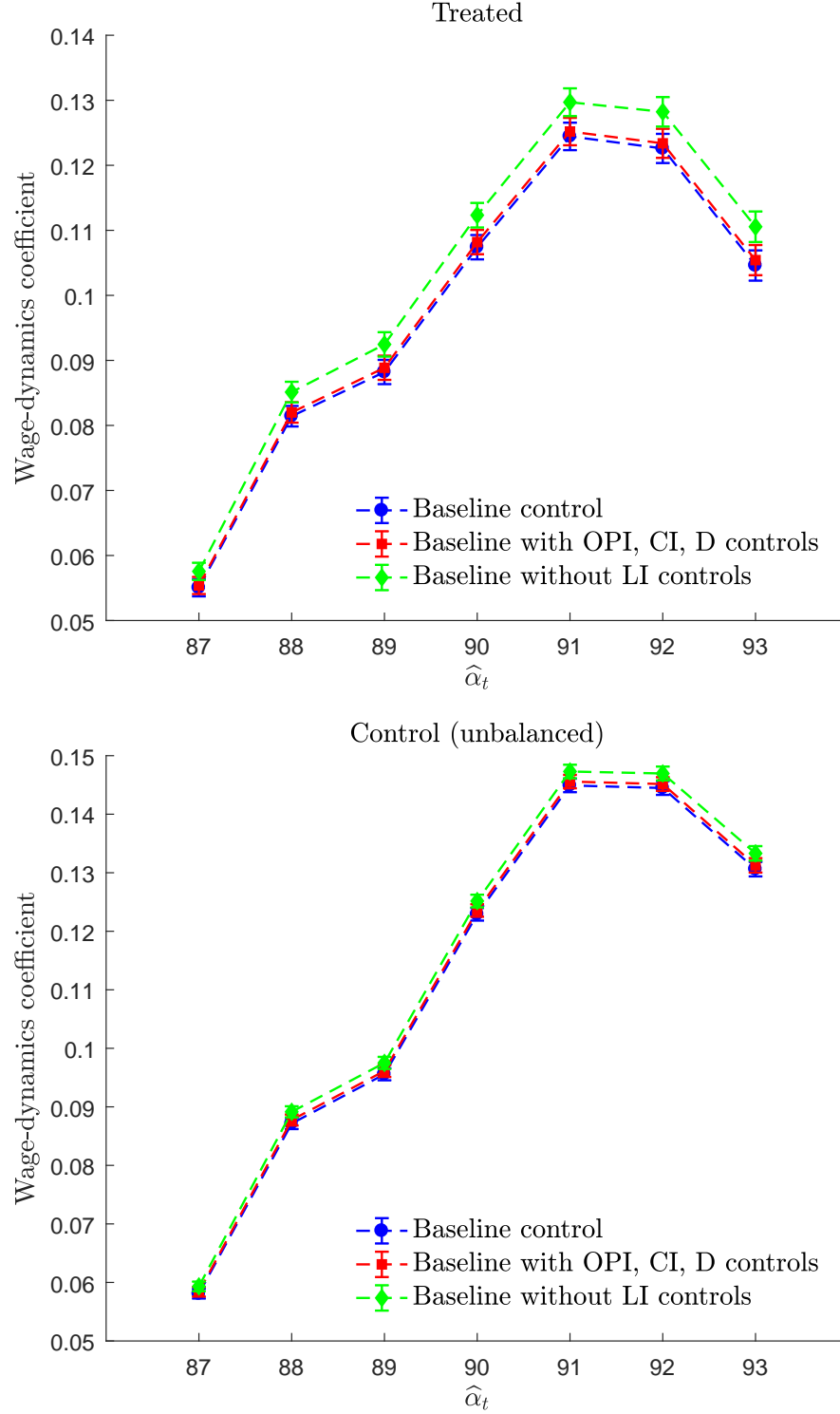
Figure 1.8: Wage response



Notes: The figure plots point estimates of wage elasticities with their 95% confidence intervals. Wage elasticities are computed as  $\hat{\epsilon}_t^w := \hat{\beta}_t^w / \hat{\delta}_1$ , where  $\hat{\delta}_1$  is from the first-stage regression specified by Equation (1.3).  $\hat{\beta}_t^w$  is from the second-stage regression, where “Treatment effect” (blue) is based on Equation (1.2) while “Placebo test” (red) is based on Equation (1.4). Standard errors in the second stage are clustered by individual. Standard errors for  $\hat{\epsilon}_t^w$  are adjusted by the delta method treating  $\hat{\delta}_1$  as a constant variable. Control units are weighted by entropy balancing.

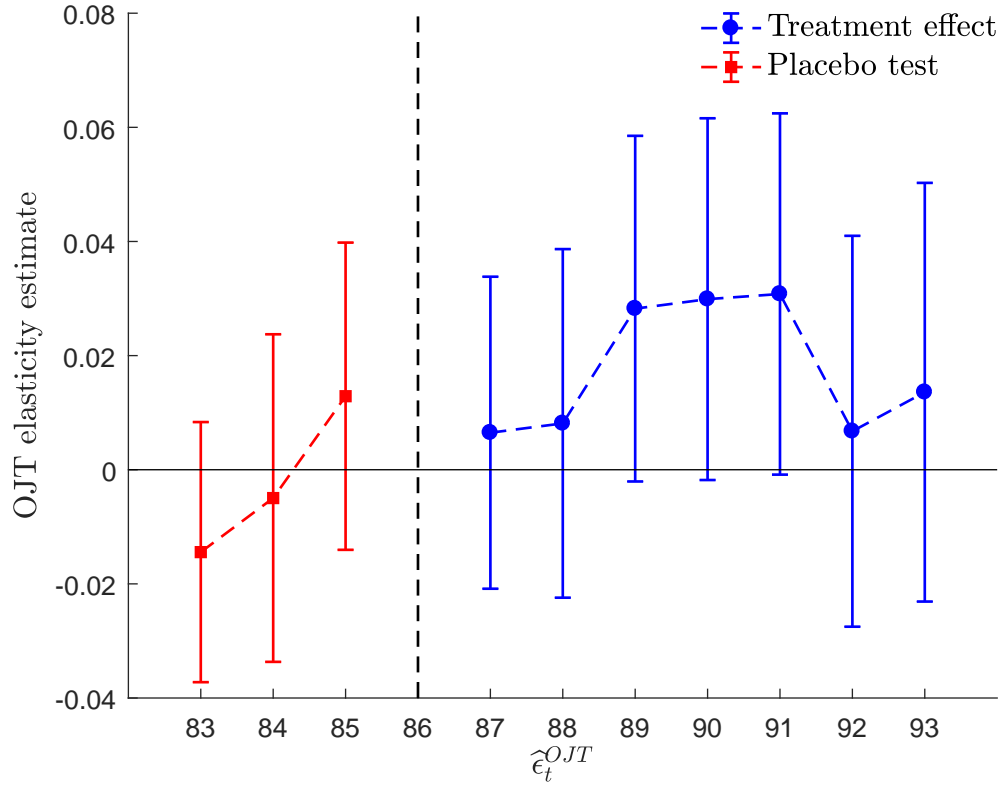


Figure 1.9: Exclusion restriction test



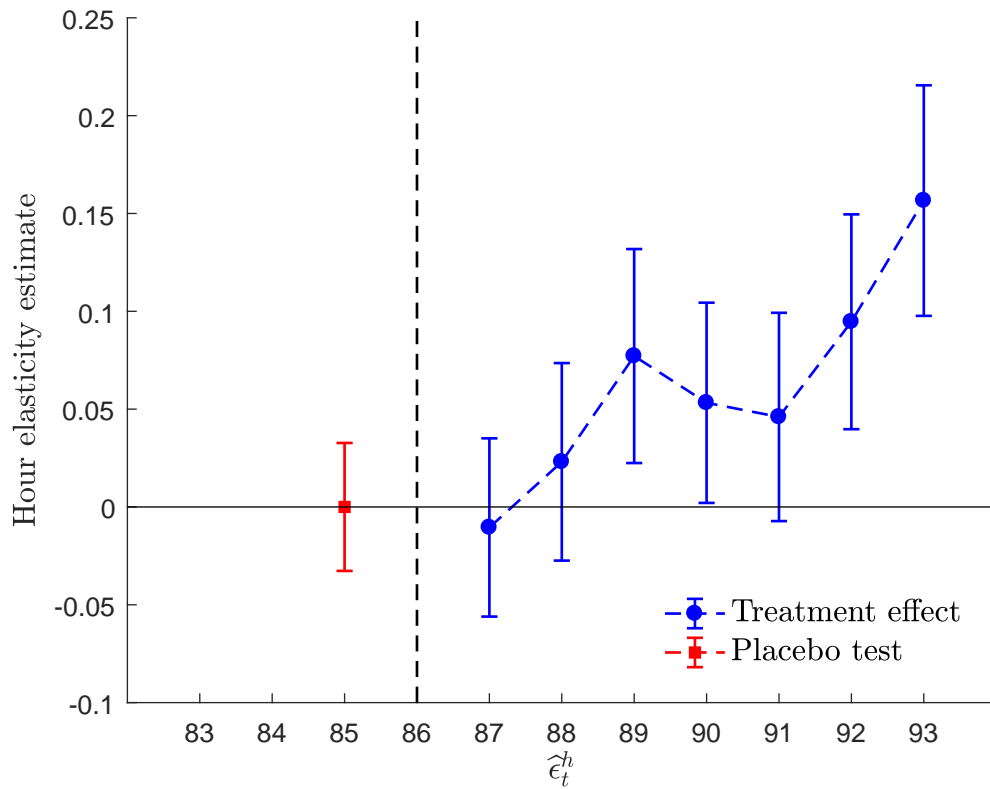
Notes: The figure plots point estimates of wage-dynamics coefficients with their 95% confidence intervals. Wage-dynamics coefficients  $\hat{\alpha}_t$  are from the regression specified by Equation (1.5). Standard errors are clustered by individual. Control units are raw and not weighted by entropy balancing.

Figure 1.10: On-the-job training response



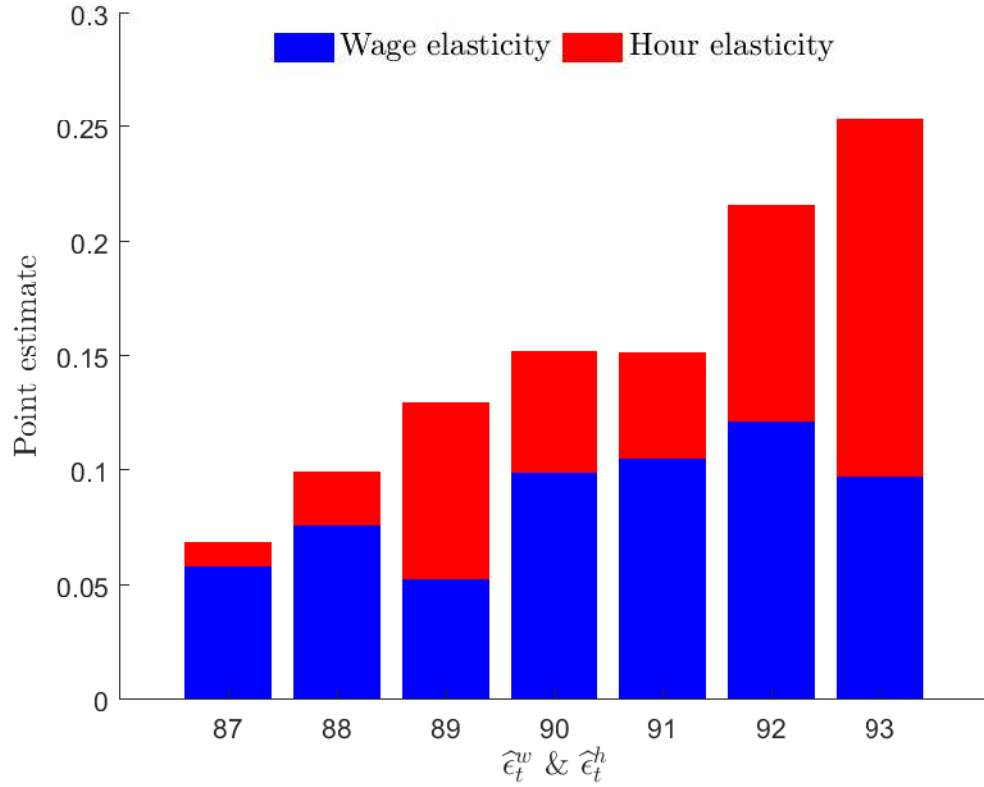
Notes: The figure plots point estimates of OJT elasticities with their 95% confidence intervals. OJT elasticities are computed as  $\hat{\epsilon}_t^{OJT} := \hat{\beta}_t^{OJT} / \hat{\delta}_1$ , where  $\hat{\delta}_1$  is from the first-stage regression specified by Equation (1.3).  $\hat{\beta}_t^{OJT}$  is from the second-stage regression, where “Treatment effect” (blue) is based on Equation (1.6) while “Placebo test” (red) is analogous to Equation (1.4) with  $1\{OJT_{it}\}$  as the dependent variable. Standard errors in the second stage are clustered by individual. Standard errors for  $\hat{\epsilon}_t^{OJT}$  are adjusted by the delta method treating  $\hat{\delta}_1$  as a constant variable. Control units are weighted by entropy balancing.

Figure 1.11: Hour response



Notes: The figure plots point estimates of hour elasticities with their 95% confidence intervals. Hour elasticities are computed as  $\hat{\epsilon}_t^h := \hat{\beta}_t^h / \hat{\delta}_1$ , where  $\hat{\delta}_1$  is from the first-stage regression specified by Equation (1.3).  $\hat{\beta}_t^h$  is from the second-stage regression, where “Treatment effect” (blue) is based on Equation (1.7) while “Placebo test” (red) is analogous to Equation (1.4) with  $\log h_{it}$  as the dependent variable. Standard errors in the second stage are clustered by individual. Standard errors for  $\hat{\epsilon}_t^h$  are adjusted by the delta method treating  $\hat{\delta}_1$  as a constant variable. Control units are weighted by entropy balancing.

Figure 1.12: Welfare implication of wage and hour responses



Notes: The figure stacks point estimates of wage and hour elasticities. Wage elasticities are computed as  $\hat{\epsilon}_t^w := \hat{\beta}_t^w / \hat{\delta}_1$ , where  $\hat{\delta}_1$  is from the first-stage regression specified by Equation (1.3).  $\hat{\beta}_t^w$  is from the second-stage regression based on Equation (1.2). Hour elasticities are computed as  $\hat{\epsilon}_t^h := \hat{\beta}_t^h / \hat{\delta}_1$ .  $\hat{\beta}_t^h$  is from the second-stage regression based on Equation (1.7). Control units are weighted by entropy balancing. Since the point estimate of  $\hat{\epsilon}_{87}^h$  is negative, its absolute value is stacked in the figure.

## Chapter 2

# Income Taxation and the Equilibrium Allocation of Labor

### 2.1 Introduction

Labor income taxation is a key vehicle by which governments redistribute income, alleviate risk and finance public goods and services. Indeed, developed economies raise 35%-50% of national income in taxes, with labor income taxation being the primary source of revenue, accounting for roughly 75% of the total tax burden (Piketty and Saez, 2013). However, at the same time, taxation distorts incentives, leaves gains from trade unexploited, and hampers economic efficiency and growth. The design of optimal income tax schedules that achieve redistributive goals with minimum distortionary effects requires detailed knowledge of how economic agents' respond to taxation. As a result, a large and diverse empirical literature studying behavioral responses to income taxation has emerged.<sup>1</sup> We contribute to this literature by studying the impact of labor income taxation on workers' job search behavior and the implications it has for the equilibrium allocation of heterogeneous workers across heterogeneous firms.

Our analysis is based on a rich equilibrium on-the-job search model with two-sided heterogeneity, endogenous job search effort on the worker-side, as in Burdett (1978), Christensen, Lentz, Mortensen, Neumann, and Werwatz (2005), and Bagger and Lentz (2015), endogenous hiring intensity on the firm-side, wage determination as in Burdett and Mortensen (1998) and Bontemps, Robin, and van den Berg (2000), and firm entry and exit as in Melitz (2003). Curvature in the utility of consumption implies that workers' optimal search effort choices feature both substitution and income effects, and introduces a desire

---

<sup>1</sup>See e.g. the Mirrlees Review (Adam, Besley, Blundell, Bond, Chote, Johnson, Myles, and Poterba, 2011; Mirrlees, Adam, Besley, Blundell, Bond, Chote, Johnson, Myles, and Poterba, 2011) for a recent and comprehensive collection of studies of issues related to taxation, including labor income taxation.

for consumption smoothing in the face of labor income risk stemming from job finding and job destruction events. We assume the presence of a complete system of insurance markets in which workers may trade securities to alleviate this consumption risk, and solve for the complete markets equilibrium using a large household formulation of the model in the mould of Merz (1995). In a nutshell, the model stipulates that income taxation reduces the return to job search effort, thus distorting workers' job search effort, and therefore the equilibrium allocation of heterogeneous workers across heterogeneous firms.

The empirical analysis makes use of a comprehensive population-wide Danish matched-employer employee panel covering 1990-2005. The data includes detailed information on workers individual tax liabilities, which allow us to obtain precise estimates of the changing Danish tax system during the observation period. We structurally estimate the equilibrium model by way of Indirect Inference using data pertaining to 1999-2003. The estimated model is used for three exercises that illustrate and quantify the impact of labor income taxation on the equilibrium allocation of labor. First, we evaluate how three Danish income tax reforms during 1990-2005 affected the steady state unemployment rate and taxable labor income. Second, we provide new estimates of a mobility-based long-run elasticity of taxable labor income, including a novel structural decomposition of the elasticity of taxable labor income. Third, we characterize a Pareto optimal income tax reform, which increases government revenue while leaving all workers (weakly) better off vis-a-vis the baseline 1999-2003 income tax regime in terms of steady state expected utility.

We find that the 1994, 1999, and 2004 Danish income tax reforms improved the equilibrium allocation of labor, reducing the steady state unemployment rate by about 6 percent, and increasing steady state labor income by about 2 percent, with workers of low ability gaining the most in terms of lower unemployment rates and higher labor incomes. Currently, our estimation procedure and subsequent quantitative analysis does not balance the government budget, and overall, the tax reforms during 1990-2005, lead to a 2 percent drop in government revenue. Accounting for equilibrium adjustments is quantitatively important when evaluating the three income tax reforms.

The estimated model implies that the full equilibrium mobility-based elasticity of taxable labor income is 0.12. The corresponding compensated elasticity of taxable labor income, where any income effects have been eliminated, is larger, at 0.17. Moreover, we find that workers of lower ability have higher taxable labor income elasticities. A structural decomposition reveal that 46 percent of the measured elasticity of taxable income come from extensive margin allocative adjustments (i.e. changes in the steady state unemployment rate), 12 percent come from intensive margin allocative adjustments (reflecting changes in the steady state distribution of employed workers across firms), and 42 percent come from firms' wage policy adjustments. When we apply the decomposition across the

ability distribution, we see that extensive margin adjustments drives the bulk of the measured taxable labor income elasticity at the bottom of the ability distribution, whereas intensive margin adjustments and adjustment to wage policies drive the taxable labor income elasticity estimates towards the top of the ability distribution.

Finally, we identify a Pareto optimal income tax reform, see Werning (2007), Blundell and Shephard (2012) and Hosseini and Shourideh (2017), which increases government revenue by 0.68 percent while leaving no workers worse off compared to the baseline 1999-2003 income tax schedule. The Pareto optimal income tax schedule reduces the marginal tax rate at low wage levels, and increases it slightly at all higher wage levels. At the same time, the average tax rate is increased at low wages, and reduced at higher wage levels. Interestingly, the Pareto optimal tax schedule shares some features of the actual tax reforms implemented in Denmark during the 1990-2005 period.

There are only a few other papers providing empirical evidence on the effect of income taxation on job search effort, labor mobility and labor allocation. Gentry and Hubbard (2004a) presents a reduced form regression analysis of the impact of income taxation on workers' labor market mobility. Using time- and state-variation in income taxation in the US as a source of identification, Gentry and Hubbard (2004a) find that a 5 percentage point reduction in the marginal tax rate increases the probability that a worker (in their study, a head of household) is in better job within a year by 0.79 percentage points, and that a reduction in income tax progressivity increases the probability that a worker moves to better job. In a paper closely related to ours, Kreiner, Munch, and Whitta-Jacobsen (2015) estimate a partial equilibrium on-the-job search model on Danish data in order to compute a mobility-based elasticity of taxable labor income. Kreiner et al. (2015) finds a taxable labor income elasticity of 0.30, but their analysis does not account for income effects in job search behavior, does not explicitly account for both worker and firm heterogeneity, and does not allow equilibrium adjustments to changes in the income tax system. Shephard (2017) develops an equilibrium search model where firm's differ in productivity, and workers differ in the value of non-market time. The model, which also features both part- and full-time jobs, is estimated using UK data, including a detailed accounting for income taxation, and is then used to conduct an equilibrium evaluation of the labor supply responses to a reform of the Working Families Tax Credit, a UK in-work benefit scheme. Shephard (2017) does not have endogenous search on the worker side and focus instead on the labor supply responses at the part-time and full-time margin, and does not use matched employer-employee data in the estimation. Breda, Haywood, and Wang (2017) develops an equilibrium search model to analyze labor market effects of taxation and minimum wage policies in France, but do not endogenize on-the-job search intensity, and thus primarily focus on workers' incentive to leave unemployment.

The paper proceeds as follows. Section 2.2 presents the model and characterizes the equilibrium, section 2.3 describes the data we use, section 2.4 provides a detailed description of the Danish income tax system during the relevant 1990-2005 period and also describes how we estimate the tax function used in the estimation. Section 2.5 details the estimation procedure, including a discussion of the estimated model's fit to data, and a presentation of the estimated structural parameters, section 2.6 contains the equilibrium evaluation of the 1994, 1999, and 2004 Danish income tax reforms, section 2.7 contains our analysis of the elasticity of taxable income, and section 2.8 presents a Pareto optimal income tax reform based on the estimated model. Finally, section 2.9 concludes, and a number of appendices contain technical details and supplementary results.

## 2.2 The model

### 2.2.1 The environment

The model is set in continuous time and the future is discounted at rate  $\rho$ . We assume throughout that the labor market is in steady state.

**Two-sided heterogeneity.** One side of the labor market is populated by a unit measure of infinitely lived individuals of heterogeneous scalar ability  $a \in [0, 1]$ . Let  $H(\cdot)$  be the CDF of ability in the worker population. We take  $H(\cdot)$  to be continuously differentiable with PDF  $h(\cdot)$ . On the other side of the market, there is a measure  $M_0$  firms of heterogeneous productivity  $p \in [p_0, 1]$ . Productivity  $p$  is distributed with CDF  $\Gamma_0(\cdot)$  across firms.  $M_0$ ,  $p_0$  and  $\Gamma_0(\cdot)$  are endogenous objects.

**Match output.** A match between an ability- $a$  worker and a productivity- $p$  firm yields output  $y(a, p)$ . The match production function  $y(a, p)$  is assumed twice continuously differentiable with  $y'_a(a, p) > 0$  and  $y'_p(a, p) > 0$  for all  $(a, p)$ -combinations. All other derivatives of the match output function are left unrestricted. The total output of a firm is the sum of the output from the matches it participates in. That is, the production function at the firm-level exhibits constant returns to scale in matches.

**Match formation and dissolution.** We assume labor markets are segregated by worker-ability  $a$ . Search frictions require workers and firms to exert effort to form productive matches, and workers may search both off- and on-the-job. A type- $a$  worker who exerts job search effort  $s$  receives job offers from hiring firms at rate  $\lambda(a)s$ , where  $\lambda(a)$  is the job finding rate per unit of search effort for type- $a$  workers, which is an equilibrium object that depends on the tightness of the market for type- $a$  labor.



A job offer is a draw of a gross wage  $w$  from the ability-conditional wage offer distribution  $F(w|a)$  with support  $[\underline{w}(a), \bar{w}(a)]$ . When a worker accepts a job offer, a match is formed, and the worker starts working for the new employer immediately. Matches are dissolved in one of three ways. First, workers are laid off at rate  $\delta(a)$  upon which they transition into unemployment. Second, workers quit for higher paying job found through on-the-job search. This event occurs at rate  $\lambda(a)s\bar{F}(w|a)$ , where  $s$  is the worker's endogenous search effort, and where  $\bar{F}(w|a) \equiv 1 - F(w|a)$ . Third, workers are exogenously reallocated to new matches at rate  $\lambda(a)\mu$ , with a wage drawn from  $F(w|a)$ . The worker can reject a reallocation and instead transition to unemployment. Descriptively, the reallocation shocks generates transitions from high to low paying jobs, events that are empirically prevalent, but that the model would otherwise be unable to generate.<sup>2</sup> To simplify the model solution, unemployed workers are also subject to reallocation shocks.

**Taxation and the government's budget.** A worker's net wage resulting from a gross wage of  $w$  is  $w - T(w)$ , where  $T(\cdot)$  is the income tax function. Taxes are paid continuously, and we assume that  $T(w) < w$ , that  $T(\cdot)$  is twice differentiable, and that  $0 \leq T'(w) < 1$ . Firm profit is taxed at rate  $\tau$ . An unmatched worker receives gross unemployment benefit  $b$ , and the net flow income of an unemployed worker is  $b - T(b)$ . If the government's steady state revenue from levying income and profit taxes is  $R$ , the government budget is the identity

$$B = R - n^0 b - \bar{b} \equiv 0, \quad (2.1)$$

where, anticipating some notation to be introduced below,  $n^0$  is the aggregate steady state unemployment rate, and  $\bar{b}$  is a per-capita lump-sum transfer that balance the government budget.<sup>3</sup>

**Worker preferences.** Workers have preferences over consumption of a composite private good  $c$  and job search effort  $s$ , an economic bad that yields disutility. The preferences are represented by the additively separable felicity function  $\psi : [0, \infty) \times [\underline{s}, \infty) \mapsto \mathbb{R}$ , with

$$\psi(c, s) = u(c) - \zeta(s - \underline{s}), \quad (2.2)$$

where  $u : [0, \infty) \mapsto \mathbb{R}$  is a continuously differentiable function with  $u'(\cdot) > 0$  and  $u''(\cdot) \leq 0$ , and  $\zeta : [0, \infty) \mapsto [0, \infty)$  is a continuously differentiable function with  $\zeta'(\cdot) > 0$  and

---

<sup>2</sup>In terms of interpretation, transitions from high to low paying jobs may reflect advanced layoff notices, strong duration dependence in the unemployment-to-job hazard, or simply considerations that fall outside the scope of the model, e.g. amenities.

<sup>3</sup>As will become clear as we present the model, imposing a balanced budget is not a trivial exercise and it involves some numerical complications that we have yet to solve in a satisfactory way. For this reason, our current estimation procedure and analysis is done with  $\bar{b} = 0$ .

$\zeta''(\cdot) > 0$ . This parameterization bounds search effort from below by  $\underline{s} > 0$ , reflecting that low levels of search effort, up until  $s = \underline{s}$ , involves no disutility. We refer to  $\underline{s}$  as “free search”. Free search turns out not to be quantitatively important, but rules out degenerate equilibrium wage offer distributions (Diamond, 1971). As we shall see, the convexity of  $\zeta(\cdot)$  determines the substitution effect in the job search response to income tax reforms. The income effect is pinned down by the concavity of  $u(\cdot)$ . Insofar as the job search literature has endogenized search effort, it typically imposes quasi-linear preferences that rules out income effects in job search behavior.<sup>4</sup> With quasi-linear preferences, a reduction in the marginal income tax rate invariably result in increased worker search effort. In the more general case of (2.2), search effort may be reduced if the income effect dominates the substitution effect.

**Large households and the complete markets allocation.** Curvature in the utility of consumption,  $u(\cdot)$ , also introduces a desire for consumption smoothing in the face of labor income risk stemming from job finding and job destruction events. We deal with this issue within the complete markets paradigm, and follow Merz (1995) in characterizing the resulting allocation (of consumption and search effort) by introducing a set of large households in the economy. Specifically, each worker in the economy belong to large household that is able to dictate the job search effort of its members. Members of a household pool their income to provide perfect consumption insurance against labor income fluctuations. A household consists of a unit measure of workers of identical ability. We refer to a household of ability- $a$  workers as a type- $a$  household. There are  $h(a)$  type- $a$  households, coinciding with the measure of ability- $a$  workers, and the economy thus contains a unit measure of households,  $\int_0^1 h(a)da = 1$ .

Even though our analysis is strictly positive—measuring and analyzing the consequences of the actual tax schedule in place and side-stepping any normative consideration regarding the shape of this tax schedule—we note here that the complete markets assumption eliminates the need for government-provided social insurance through progressive taxation. However, as our model features ex-ante heterogeneous workers, a progressive income tax schedule may nonetheless be warranted out of concerns for redistribution. Heathcote, Storesletten, and Violante (2017) provide a thorough analysis of the factors that shape the optimal degree of income tax progressivity. Michau (2017) analyze optimal provision of social insurance in an on-the-job search model of ex-ante homogenous workers.

---

<sup>4</sup>Indeed, existing empirical work on the effect of income taxation on job search and job mobility either takes no account of income effects (see Kreiner et al., 2015), or does not separate income and substitution effects (Gentry and Hubbard, 2004a). This is in sharp contrast to the empirical labor supply literature, see e.g. Heckman (1993) and Blundell and MaCurdy (1999).

**Wage determination.** The ability-conditional wage offer distribution  $F(w|a)$  is determined in a wage posting equilibrium as in Burdett and Mortensen (1998) and Bontemps et al. (2000). The wage setting game restricts admissible wage policies in three dimensions. First, firms make take-it-or-leave-it offers and credibly commit to future wage payments. Second, the wage must remain constant throughout the duration of the match. Third, contracts are anonymous, implying that a firm must offer the same contract to all workers of the same ability. A firm's wage policy maximizes its steady state profit by trading off profit per worker and worker turnover, and is a best response to all other firms' wage and recruitment policies and workers' optimal search effort choices. The equilibrium outcome is a set of non-degenerate ability-specific wage offer distributions, where more productive firms offer higher wages.

Alternative wage setting arrangements are available. Coles (2001) characterize a wage posting equilibrium where firms may renege on promised wage payments. Stevens (2004), and Burdett and Coles (2003) and Burdett and Coles (2010) relax the assumption that the offered wage remain constant throughout the match and characterize the set of (privately) optimal wage-tenure contracts. The contracts derived in Stevens (2004) has little empirical content. Burdett and Coles (2003) and Burdett and Coles (2010) consider risk-averse workers who inhabit an incomplete markets economy where there is value to firms smoothing workers consumption paths through the wage contract. While the resulting wage contracts are empirically plausible, with complete markets, there is no need for firms to provide such insurance. Postel-Vinay and Robin (2002) and Cahuc, Postel-Vinay, and Robin (2006) relax the assumption of take-it-or-leave-it offers through the introduction of sequential auctions and bargaining in the presence of on-the-job search.<sup>5</sup> Lentz (2010) and Bagger and Lentz (2015) extend the Cahuc et al. (2006) framework to allow for endogenous job search effort. However, with income and profit taxation, wages are no longer pure transfers in the bargaining game, rendering the Cahuc et al. (2006) framework unattractive for our purposes.<sup>6</sup> Finally, Lentz (2015) studies equilibrium wage formation in an incomplete markets economy where on-the-job search effort is subject to moral hazard.

---

<sup>5</sup>From a theoretical point of view, Postel-Vinay and Robin (2004) consider whether firms should match outside offers and reach the (tentative) conclusion that more productive firms may benefit from an offer-matching strategy, whereas firms with lower productivity may be better off committing not to match outside offers. From an empirical point of view, using survey data, Hall and Krueger (2012) find that about 1/3 of the workers in their survey have wages set through take-it-or-leave-it offers, whereas another 1/3 were able to bargain.

<sup>6</sup>Breda et al. (2017) uses a sequential bargaining model in the mould of Cahuc et al. (2006) to analyze the effect of payroll taxes on labor allocation, but does not provide a solution to the wage bargaining game, instead imposing a reduced form sharing rule.

### 2.2.2 Household behavior

Consider a type- $a$  household. Let  $n_t^0(a)$  be the measure of unemployed household members at time  $t$ , and let  $n_t^1(w|a)$  be the measure of household members employed with a gross wage  $w$  at time  $t$ . Define

$$N_t(a) \equiv n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} n_t^1(w|a)dw \quad (2.3)$$

as the time- $t$  measure of the household population.

The distribution of workers across unemployment and employment, and within employment, across wage levels depend on household members' labor market behavior. The household is able to dictate, and enforce, the search effort and acceptance decisions of its members. Let  $s_t^0(a)$  be the dictated time- $t$  search intensity for an unemployed type- $a$  worker, and let  $s_t^1(w|a)$  be the dictated time- $t$  search intensity for a type- $a$  worker with gross wage  $w$ . In regard to the dictated acceptance decision of job offers, we confirm below that the household will employ a reservation wage strategy. Let  $\phi(a)$  be the reservation wage dictated to unemployed workers in a type- $a$  household. The reservation wage of members employed with wage  $w$  is, obviously,  $w$ .

Members of the household pool their net-of-tax labor income and perfectly insure each other against consumption fluctuations due to individual job finding and job destruction shocks. Hence, time- $t$  consumption of a type- $a$  household member is independent of current individual labor market status and labor income, and is given by

$$c_t(a) = \frac{1}{N_t(a)} \left( [b - T(b)]n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} [w - T(w)]n_t^1(w|a)dw \right) + \bar{b}, \quad (2.4)$$

the average type- $a$  household net-of-tax income at time  $t$ . Recall that  $\bar{b}$  is a government lump-sum transfer that balances the government budget.

Turning to the labor market behavior, a type- $a$  household set unemployed search effort profile  $s_t^0(a)$ , an employed search effort profile  $s_t^1(w|a)$  for every  $w \in [\phi(a), \bar{w}(a)]$ , as well as a reservation wage  $\phi(a)$ , in order to maximize the present value household lifetime utility,  $\Psi(a)$ , given as

$$\Psi(a) = \int_0^\infty e^{-\rho t} \left[ \psi(c_t(a), s_t^0(a))n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} \psi(c_t(a), s_t^1(w|a))n_t^1(w|a)dw \right] dt, \quad (2.5)$$

where  $\psi(c, s)$  is the felicity function given by (2.2), and  $c_t(a)$  is the household consumption level given by (2.4). The optimization is subject to the law of motions for  $n_t^0(a)$  and

$n_t^1(w|a)$ . The law of motion for  $n_t^0(a)$  is given by

$$\dot{n}_t^0(a) = \delta(a)N_t(a) - [\delta(a) + \lambda(a)\mu + \lambda(a)s_t^0(a)]n_t^0(a), \quad (2.6)$$

The “dot”-notation indicates derivatives with respect to time  $t$ . The law of motion for  $n_t^1(w|a)$  is slightly more involved and is given by

$$\begin{aligned} \dot{n}_t^1(w|a) = & \left[ \lambda(a)\mu N_t(a) + \lambda(a)s_t^0(a)n_t^0(a) + \int_{\phi(a)}^w \lambda(a)s_t^1(x|a)n_t^1(x|a)dx \right] f(w|a) \\ & - \left[ \delta(a) + \lambda(a)\mu + \lambda(a)s_t^1(w|a)\bar{F}(w|a) \right] n_t^1(w|a), \quad (2.7) \end{aligned}$$

for every  $w \in [\phi(a), \bar{w}(a)]$ . The derivation of both (2.6) and (2.7) are based on gross worker flow accounting and are detailed in Appendix A.1.1.

Conditional on  $\phi(a) < \bar{w}(a)$ , the household problem is an optimal control problem with the controls  $s_t^0(a)$  and  $s_t^1(w|a)$  for every  $w \in [\phi(a), \bar{w}(a)]$ , and states  $n_t^0(a)$  and  $n_t^1(w|a)$  for every  $w \in [\phi(a), \bar{w}(a)]$ . Define  $\xi_t^0(a)$  as the multiplier on the law of motion for  $n_t^0(a)$  in the household’s problem, and let  $\xi_t^1(w|a)$  be the multiplier on the law of motion for  $n_t^1(w|a)$ .  $\xi_t^0(a)$  and  $\xi_t^1(w|a)$  are referred to as the costate variables, and we denominate them in present ( $t = 0$ ) values. The costates can be interpreted as shadow values, and are central to the analysis to come. Indeed,  $\xi_t^0(a)$  is the increase in the present value household lifetime utility  $\Psi(a)$ , see (2.5), from adding an additional unemployed worker to the household at time  $t$ . Likewise,  $\xi_t^1(w|a)$  is the increase in  $\Psi(a)$  from adding an extra worker with wage  $w$  to the household at time  $t$ .

Our analysis focuses on an economy in steady state, where controls, states and the *current value* costates, given by  $e^{\rho t}\xi_t^0(a)$  and  $e^{\rho t}\xi_t^1(w|a)$ , are time-invariant, and the *present value* costates,  $\xi_t^0(a)$  and  $\xi_t^1(w|a)$ , varies with  $t$  only because of discounting. That is, steady state behavior is captured by  $\xi_0^0(a)$  and  $\xi_0^1(w|a)$  with the following set of steady state identities applying:  $\xi_0^0(a) = e^{\rho t}\xi_t^0(a)$  and  $\xi_0^1(w|a) = e^{\rho t}\xi_t^1(w|a)$  for every  $w \in [\phi(a), \bar{w}(a)]$ .<sup>7</sup> Steady state values of  $s_t^0(a)$  and  $s_t^1(w|a)$  are denoted by  $s^0(a)$  and  $s^1(w|a)$ , respectively.

Suppose an unemployed household member has received an offer of a job paying a wage  $w$ . In order to maximize  $\Psi(a)$ , the household dictates a rejection if the value of the worker remaining unemployed,  $\xi_0^0(a)$ , exceeds the value of having the worker employed in a job paying  $w$ ,  $\xi_0^1(w|a)$ . The household dictates acceptance if the opposite inequality between  $\xi_0^0(a)$  and  $\xi_0^1(w|a)$  holds. We confirm below that  $\xi_0^1(w|a)$  is strictly increasing in  $w$ . Hence, the household employs a reservation wage strategy, and the current reservation

---

<sup>7</sup>  $\xi_0^0(a)$  and  $\xi_0^1(w|a)$  are the added value to  $\Psi(a)$  from immediately admitting an additional unemployed worker, respectively employed worker with wage  $w$ , to the household when the economy is in steady state.

wage  $\phi(a)$  solves

$$\xi_0^1(\phi(a)|a) = \xi_0^0(a). \quad (2.8)$$

The household's optimal control problem can be solved by constructing the associated Hamiltonian and applying a version of the Maximum Principle (Acemoglu, 2009, section 7.5, p. 253). We state just the main results here, with details on the derivations relegated to Appendix A.1.3. Optimal steady state search effort for employed workers earning a wage  $w \in [\phi(a), \bar{w}(a)]$  satisfies

$$\zeta'(s^1(w|a) - \underline{s}) = \lambda(a) \int_w^{\bar{w}(a)} [\xi_0^1(x|a) - \xi_0^1(w|a)] dF(x|a) = \lambda(a) \int_w^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_0^1(x|a) \bar{F}(x|a) dx, \quad (2.9)$$

where the last equality follows from integrating by parts. The interpretation of (2.9) is straightforward. Suppose the household perturb its steady state search policy by dictating marginally higher search intensity at wage-rung  $w$ , specifically from  $s^1(w|a)$  to  $s^1(w|a) + \Delta$ . This perturbation involves an additional disutility-flow  $\zeta'(s^1(w|a) - \underline{s})\Delta$  per rung- $w$  worker, see (2.2). However, it also increases the rate at which rung- $w$  workers obtain job offers by  $\lambda(a)\Delta$ . The additional offers are distributed according to  $F(\cdot|a)$ , but only a share  $\bar{F}(w|a)$  are acceptable. Reallocating a worker from rung  $w$  to rung  $x > w$  yields a household utility gain of  $\xi_0^1(x|a) - \xi_0^1(w|a)$ . Hence, the expected per-worker marginal value gain of the search policy perturbation is  $\lambda(a)\Delta \int_w^{\bar{w}(a)} [\xi_0^1(x|a) - \xi_0^1(w|a)] dF(x|a)$ . Condition (2.9) therefore balances the added marginal disutility from search with the expected marginal gains from search. An analogous condition characterizes optimal unemployed search.<sup>8</sup>

A recursive expression for the steady state value to the household of a worker employed at wage  $w$ ,  $\xi_0^1(w|a)$ , can be derived as

$$\begin{aligned} \rho \xi_0^1(w|a) = & \psi(c(a), s^1(w|a)) + u'(c(a))[w - T(w) - c(a)] \\ & + \delta(a)[\xi_0^0(a) - \xi_0^1(w|a)] + \lambda(a)\mu \int_{\phi(a)}^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_0^1(x|a) \bar{F}(x|a) dx \\ & + \lambda(a)s^1(w|a) \int_w^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_0^1(x|a) \bar{F}(x|a) dx, \end{aligned} \quad (2.10)$$

where  $\psi(\cdot, \cdot)$  is the felicity function given by (2.2), and where  $c(a)$  is the steady state version of household consumption, see (2.4). A worker employed with wage  $w$  is an asset to the household. The left-hand side of (2.10) is the permanent utility flow from this asset. This flow comprises the household payoff flow, and the expected returns to job destruction, job reallocation, and job finding. The marginal wage-rung  $w$  worker

---

<sup>8</sup>Optimal steady state unemployed search effort  $s^0(a)$  solves  $\zeta'(s^0(a) - \underline{s}) = \lambda(a) \int_{\phi(a)}^{\bar{w}(a)} [\xi_0^1(x|a) - \xi_0^0(a)] dF(x|a)$ .

increases the household utility flow by worker-felicity  $\psi(c(a), s^1(w|a))$ , which leaves net-of-tax revenue in the amount of  $w - T(w) - c(a)$  to be distributed within the household. The household values this additional revenue at the marginal utility of consumption,  $u'(c(a))$ . Job destruction, job reallocation, and job finding events are associated with capital gains from the worker transitioning into unemployment, being reallocated to a randomly drawn new job, or being matched with a new potential employer through on-the-job search.

In Appendix A.1.3 we derive a similar expression for the steady state value of an unemployed worker,  $\xi_0^0(a)$ , which reads

$$\begin{aligned} \rho \xi_0^0(a) = & u(c(a)) - \zeta(s^0(a)) + u'(c(a))[b - T(b) - c(a)] \\ & + \lambda(a)[\mu + s^0(a)] \int_{\phi(a)}^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_0^1(x|a) \bar{F}(x|a) dx. \end{aligned} \quad (2.11)$$

Substituting (2.10) and (2.11) into the reservation wage definition (2.8) shows that

$$\phi(a) = b. \quad (2.12)$$

This result stems from the assumption that unemployed and employed job search is equally efficient, and involves equal disutility. In this case, there are no opportunity cost of taking up a job. We impose  $\phi(a) = b$  in the remainder of the paper. Given (2.12), it is evident that optimal off-the-job search equals optimal on-the-job search in a firm offering  $w = b$ , that is,  $s^0(a) = s^1(b|a)$ .<sup>9</sup> This result allow us to simplify notation by dropping the superscript “0” and “1” that has distinguished unemployed and employed search up until now. In the remainder of the paper we let  $s(w|a)$  denote optimal search effort at wage  $w$ , with  $s(b|a)$  being optimal unemployed search.

Taking the derivative of (2.10) with respect to  $w$ , applying the Envelope Theorem, and substituting the resulting expression for  $\partial \xi_0^1(w|a)/\partial w$  into (2.9) yields the following expression for the complete markets optimal job search effort as a function of the current wage  $w$ ,

$$\zeta'(s(w|a) - \underline{s}) = \lambda(a) \int_w^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)] \bar{F}(x|a)}{\rho + \delta(a) + \lambda(a)\mu + \lambda(a)s(x|a)\bar{F}(x|a)} dx. \quad (2.13)$$

The optimal search effort balances the marginal cost with the marginal gains from on-the-job search. In general, workers of different ability differ in their chose job search effort, due to differences in the return to job search stemming from differences in the distribution of job offers  $\bar{F}(x|a)$  across workers of different abilities. Lemma 1 summarizes some useful properties of the optimal search effort.

---

<sup>9</sup>For details, consult Appendix A.1.3.

**Lemma 1.** *Optimal search effort  $s(w|a)$  has the following properties (i)  $s(w|a) \geq \underline{s}$ , (ii)  $s(\bar{w}(a)|a) = \underline{s}$ , and (iii)  $\partial s(w|a)/\partial w < 0$  for every  $w \in [b, \bar{w}(a)]$ .*

*Proof.* See Appendix A.1.4. □

Lemma 1 states that job search effort is bounded from below by  $\underline{s} > 0$ . This is a useful feature of the model as it rules out degenerate equilibrium wage offer distributions, see Diamond (1971), Burdett and Judd (1983), and Burdett and Mortensen (1998). The return to job search effort diminishes at higher wages because there are only limited possibilities to climb further up the wage ladder. Indeed, at the highest paying job, the returns to on-the-job search is zero, and no job search effort is forthcoming. The model implies that unemployed workers exert more job search effort than employed workers, despite facing the same disutility of effort and the same search efficiency. The impact of income taxation on job search is discussed further below when we consider partial equilibrium substitution and income effects in the provision of job search effort. We note, however, that (2.13) implies that search effort is distorted by income taxation, and it is the average tax (i.e. the integrated marginal tax rate) that matters for distortions.<sup>10</sup> Furthermore, income taxation may have different quantitative implications for the job search behavior of workers of different abilities. These effects of income taxation will all be born out in the empirical analysis to come.

### 2.2.2.1 The complete markets steady state labor allocation

Normalizing the steady state household population  $N_t(a) \equiv 1$  allow us to interpret  $n^0(a)$  as the steady state unemployment rate among type- $a$  workers, and to define the steady state PDF of employed workers across wage rates as  $g(w|a) \equiv n^1(w|a)/[1 - n^0(a)]$ , with associated CDF  $G(w|a) = \int_b^{\bar{w}} g(w|a)dw$ . Together,  $n^0(a)$  and  $G(w|a)$  characterizes the steady state allocation of type- $a$  labor.

Using the steady state version of (2.6) with  $N_t(a) \equiv 1$  implies that

$$n^0(a) = \frac{\delta(a)}{\delta(a) + \lambda(a)\mu + \lambda(a)s(b|a)}, \quad (2.14)$$

where  $s(b|a)$  is given by (2.13). Furthermore, the steady state version of (2.6) with  $\phi(a) = b$

---

<sup>10</sup>Integration by parts of (2.13) results in a differential equation for  $s(w|a)$  that does not depend on the marginal tax function  $T'(\cdot)$ , but only the average tax function  $T(\cdot)$ .



and  $N_t(a) \equiv 1$  yields the following balanced flow equation defining  $G(w|a)$ ,

$$\begin{aligned} [\lambda(a)\mu + \lambda(a)s^0(a)]F(w|a)\frac{n^0(a)}{1-n^0(a)} + \lambda(a)\mu F(w|a)\overline{G}(w|a) \\ = \delta(a)G(w|a) + \overline{F}(w|a) \int_b^w \lambda(a)s(x|a)g(x|a)dx. \end{aligned} \quad (2.15)$$

The left-hand side of reflects the inflow into the stock of workers earning a wage  $w$  or less, and the right-hand side reflects the outflow. In steady state, in- and outflow balance. We show in Appendix A.1.2 that (2.15) implies the following closed-form expression for the allocation of labor,

$$G(w|a) = 1 - \exp \left\langle - \int_b^w \left[ \frac{\delta(a) + \lambda(a)\mu}{\delta(a) + \lambda(a)\mu + \lambda(a)s(x|a)\overline{F}(x|a)} \right] \frac{f(x|a)}{\overline{F}(x|a)} dx \right\rangle. \quad (2.16)$$

Since all type- $a$  households are identical,  $G(w|a)$  represents both the allocation of labor within a type- $a$  household, as well as the economy-wide allocation of type- $a$  workers.

#### 2.2.2.2 Tax reforms and partial equilibrium substitution and income effects

The analysis to follow is based on a comparative statics analysis of steady state partial equilibrium predictions of the impact of income taxation on job search effort. That is, we compare the steady state job search policies under two different tax regimes, with no account of the out-of-steady state adjustment process.

Consider a tax reform that changes the income tax function from  $T(\cdot)$  to  $\hat{T}(\cdot)$ . Let  $s(w|a)$  be the pre-reform job search effort of a type- $a$  worker employed with wage  $w$  given by (2.13), and let  $\hat{s}(w|a)$  be the post-reform search effort defined by the post-reform version of (2.13),

$$\zeta'(\hat{s}(w|a)) = \lambda(a) \int_w^{\overline{w}(a)} \frac{u'(\hat{c}(a))[1 - \hat{T}'(x)]\overline{F}(x|a)}{\rho + \delta(a) + \lambda(a)\mu + \lambda(a)\hat{s}(x|a)\overline{F}(x|a)} dx, \quad (2.17)$$

where  $\hat{c}(a)$  is post-reform consumption, computed according to the steady state version of (2.4) imposing the post-reform tax system and the post-reform labor allocation. Finally, let  $\tilde{s}(w|a)$  be the job search intensity of a type- $a$  worker employed with wage  $w$  under the new tax regime  $\hat{T}(\cdot)$ , but holding the marginal utility of consumption constant at the pre-reform level, i.e. at  $u(c(a))$ . This counterfactual object is defined by the functional

$$\zeta'(\tilde{s}(w|a)) = \lambda(a) \int_w^{\overline{w}(a)} \frac{u'(c(a))[1 - \hat{T}'(x)]\overline{F}(x|a)}{\rho + \delta(a) + \lambda(a)\mu + \lambda(a)\tilde{s}(x|a)\overline{F}(x|a)} dx. \quad (2.18)$$

The full partial equilibrium on-the-job search effort response of a type- $a$  worker em-

ployed with wage  $w$ ,  $\hat{s}(w|a) - s(w|a)$  can be decomposed into a substitution effect and an income effect according to

$$\hat{s}(w|a) - s(w|a) = \underbrace{\tilde{s}(w|a) - s(w|a)}_{\text{Substitution effect}} + \underbrace{\hat{s}(w|a) - \tilde{s}(w|a)}_{\text{Income effect}}. \quad (2.19)$$

To fix ideas, suppose the tax reform increases the net-of-tax rate for wages in the interval  $[\hat{w}, \hat{w} + \Delta] \in [b, \bar{w}(a)]$  for some  $\Delta > 0$ , while leaving the rest of the tax function unchanged. Suppose  $w < \hat{w}$ . Then, inspection of (2.13) and (2.18) reveals that, in response to the higher net-of-tax rates in  $[\hat{w}, \hat{w} + \Delta]$ , on-the-job search effort at wage  $w < \hat{w}$  increases, i.e. the substitution effect  $\tilde{s}(w|a) - s(w|a)$  is positive. The magnitude of the substitution effect is inversely related to the convexity of the disutility from search effort, i.e.  $\zeta''(\cdot)$ . The intuition is straightforward: An increase in the net-of-tax rate at some interval  $[\hat{w}, \hat{w} + \Delta] \in (w, \bar{w}(a)]$  increases the gains from reallocating to this wage interval, mandating an increase in search effort at wage  $w$  to rebalance marginal costs and benefits. If the disutility from search is highly convex, only small adjustments in search effort are needed to equalize marginal costs and benefits. If  $w > \hat{w} + \Delta$ , the reform affect only wages below the worker's current position, leaving job search incentives unchanged, and the substitution effect is zero.<sup>11</sup>

The increased net-of-tax wages in the interval  $[\hat{w}, \hat{w} + \Delta]$  also increases workers' permanent income, and thus consumption. That is,  $c(a) < \hat{c}(a)$ , and by the convexity of the utility of consumption,  $u'(c(a)) \geq u'(\hat{c}(a))$ . Inspection of (2.17) and (2.18) reveals that the magnitude of the income effect is driven by the curvature of the utility of consumption,  $u(\cdot)$ . Again, the intuition is straightforward: An increase in the net-of-tax rate at some interval  $[\hat{w}, \hat{w} + \Delta]$  increases the worker's permanent net-of-tax income, reducing the marginal utility of consumption and thus mandating a decrease in search effort at all wage levels. Since the income effect operates through the workers' permanent income, it is present even when  $[\hat{w}, \hat{w} + \Delta] \in [b, w]$ .

### 2.2.3 Firm behavior

Firm behavior is modeled in a one-shot game. Conditional on entry, a firm sets a wage policy and a hiring intensity policy in order to maximize its expected flow of net-of-tax profit in steady state. In doing so, each firm takes the income and profit tax regimes, workers' search strategies, and the hiring and wage policies of other firms as given. Each

---

<sup>11</sup>In the case where  $w \in [\hat{w}, \hat{w} + \Delta]$  a positive substitution effect arises from the incentives to reallocate to wages in the interval  $(w, \hat{w} + \Delta]$ .

firm conditions its policies on worker ability.<sup>12</sup> Specifically, a wage policy for a generic type- $p \in [p_0, 1]$  firm is  $w(a|p) : [0, 1] \mapsto \mathbb{R}^+$ , and a hiring intensity policy is  $v(a|p) : [0, 1] \mapsto \mathbb{R}^+$ . Recall that the productivity distribution among the entering firms is  $\Gamma_0(\cdot)$ . There is an underlying exogenous distribution of potential productivities  $\Gamma(p)$  with  $p \in [0, 1]$ . There is free entry of firms with entry decisions based on expected steady state net-of-tax profit flow. The entry game pins down the measure of entering firms,  $M_0$ , the minimum productivity  $p_0$ , as well as  $\Gamma_0(\cdot)$ .

### 2.2.3.1 Profit maximization

Following Burdett and Mortensen (1998) and Bontemps, Robin, and van den Berg (1999); Bontemps et al. (2000), firms that have entered the labor market seek to maximize the expected steady state flow profit.<sup>13</sup> Let  $\ell(a|w, v)$  be the the steady state measure of type- $a$  workers that a generic (type- $p$ ) firm can expect to employ when it offers this type of workers a wage  $w$  while searching for them at intensity  $v$ . The notation makes it clear that firm productivity  $p$  plays no direct role in shaping  $\ell(a|w, v)$ , but, of course, the optimal wage and hiring policies will depend on  $p$ , because output per worker depends on  $p$ . Indeed, profit per type- $a$  worker in a type- $p$  firm, paying a wage  $w$  is  $y(a, p) - w$ . Let  $\pi(p)$  be the steady state net-of-tax profit flow resulting from a type- $p$  firms' optimal wage and hiring intensity policy. Recall that profit flows are taxed at a flat rate  $\tau$ . It follows that

$$\pi(p) = \max_{w(\cdot|p), v(\cdot|p)} (1 - \tau) \int_0^1 \left\langle [y(a, p) - w(a|p)] \ell(a|w(a|p), v(a|p)) - d(v(a|p)) \right\rangle da, \quad (2.20)$$

where  $d : [0, \infty) \mapsto [0, \infty)$  is the twice continuously differentiable hiring intensity cost function. We impose  $d'(\cdot) > 0$  and  $d''(\cdot) > 0$ . Equation (2.20) embeds a number of assumptions stated above. First, firm-level output exhibits constant returns to scale in matches. Second, hiring cost is a monetary cost that is fully deductible in relation to profit taxation (say, the cost of running a human resource department). In the current formulation of the model, the hiring costs vanishes from the economy, i.e. it does not accrue to any of the agents in the economy.<sup>14</sup> We are currently working on a version of the model where firms assign employees to either production or recruitment activities, similar to the formulation in Shimer (2010). Production workers produce output according to a production function as above, whereas recruitment workers enable the firm to attract

<sup>12</sup>This setup involves a degree of directedness in firms' search processes, eliminating cross-ability congestion externalities in wage policies.

<sup>13</sup>Burdett and Mortensen (1998) and Mortensen (2005) point out that steady-state profit is the appropriate criterion only if firms do not discount the future, namely  $\rho = 0$ .

<sup>14</sup>A contrived and empirically unsatisfactory narrative underlying such a formulation could involve firms outsourcing their hiring operation to foreign owned recruitment agencies.

more workers. In such a setup, hiring cost can be regarded literally as wage costs in a human resource department, and are clearly accruing to the (recruitment) workers. Finally, a third assumption embedded in (2.20) is that hiring intensity can be directed towards workers of a particular ability, with a firms' total hiring cost being the integrated worker-type specific hiring costs.

In order for a firm to be able to hire any type- $a$  workers at all, its wage offer must exceed the type- $a$  worker reservation wage and the firm must exert strictly positive hiring effort. That is,  $\ell(a|w, v) = 0$  if and only if  $w < b$  or  $v = 0$ . Consider a generic type- $p$  firm with type- $a$  worker wage and hiring policies represented by  $w(a|p) > b$  and  $v(a|p) > 0$ . Then,  $\ell(a|w(a|p), v(a|p))$  is pinned down by the following balanced flow equation,

$$\begin{aligned} \ell(a|w(a|p), v(a|p)) \left[ \delta(a) + \lambda(a)\mu + \lambda(a)s(w(a|p)|a)\overline{F}(w(a|p)|a) \right] = \\ \frac{v(a|p)}{V(a)} \left\langle h(a)n^0(a)[\lambda(a)\mu + \lambda(a)s(b|a)] \right. \\ \left. + h(a)[1 - n^0(a)] \left[ \lambda\mu + \lambda(a) \int_b^{w(a|p)} s(x|a)dG(x|a) \right] \right\rangle, \quad (2.21) \end{aligned}$$

where  $s(w|a)$  is given by (2.13),  $n^0(a)$  is the steady state type- $a$  worker unemployment rate (2.14),  $G(w|a)$  is given by (2.16), and where  $V(a) = M_0 \int_{p_0}^1 v(a|p)d\Gamma_0(p)$  is the integrated type- $a$  worker hiring intensity. The left-hand side of (2.21) is the measure of type- $a$  workers that leave a type- $p$  firm with type- $a$  worker wage and hiring policies  $w(a|p)$  and  $v(a|p)$ . These workers leave either because they are laid off, reallocated or because they find a higher paying job. The right-hand side of (2.21) is the measure of type- $a$  workers that are hired into the type- $p$  firm. The expression in angle brackets is the number of meetings between firms and type- $a$  workers willing to accept a job offer  $w(a|p)$ . These workers are either coming from unemployment, reallocated from other firms, or are currently working in firms paying a wage less than  $w(a|p)$ . The pool of willing workers are distributed across hiring firms in proportion to the firms' hiring intensity, i.e. according to the density  $v(a|p)/V(a)$ .

Since (2.21) implies that  $\ell(a|w(a|p), v(a|p)) = \frac{v(a|p)}{V(a)} \tilde{\ell}(a|w(a|p))h(a)$ , where

$$\tilde{\ell}(a|w(a|p)) = \frac{n^0(a)[\lambda(a)\mu + \lambda(a)s(b|a)] + [1 - n^0(a)][\lambda(a)\mu + \lambda(a) \int_b^{w(a|p)} s(x|a)dG(x|a)]}{\delta(a) + \lambda(a)\mu + \lambda(a)s(w(a|p)|a)\overline{F}(w(a|p)|a)}, \quad (2.22)$$

we can analyze a firm's wage policy  $w(a|p)$  independently of its hiring intensity policy  $v(a|p)$ . Lemma 2 states a useful and intuitive property shared by  $\ell(a|w(a|p))$  and  $\tilde{\ell}(a|w(a|p))$ .

**Lemma 2.**  $\ell(a|w)$  and  $\tilde{\ell}(a|w)$  are strictly increasing in  $w$  for  $w \in [b, \bar{w}(a)]$ .

*Proof.* See Appendix A.1.5. □

The intuition behind Lemma 2 is straightforward. By offering a higher wage, a firm simultaneously increases its ability to poach workers from other (lower paying) firms, and reduces the likelihood of its own workers being poached by other (higher paying) firms. Both effects tend to increase the steady state measure of type- $a$  workers employed in the firm.

Using (2.21) and (2.22) allow us to restate the firms profit maximization problem (2.20) as a two-step problem. Indeed,

$$\pi(p) = \max_{v(\cdot|p)} (1 - \tau) \int_0^1 \left\langle \frac{v(a|p)}{V(a)} \tilde{\pi}(a|p) h(a) - d(v(a|p)) \right\rangle da, \quad (2.23)$$

where

$$\tilde{\pi}(a|p) = \max_w \left\langle [y(a, p) - w] \tilde{\ell}(a|w) \right\rangle, \quad (2.24)$$

with  $\tilde{\ell}(a|w)$  given by (2.22). Equation (2.24) pins down the optimal wage policy  $w(\cdot|p)$ , independently of the hiring policy  $v(\cdot|p)$ . Equation (2.23) then pins down the optimal hiring policy  $v(\cdot|p)$ , given the optimal wage policy.

The fact that hiring costs are fully deductible ensures that profit taxation has no direct distortionary effects on hiring intensity, nor on wage policies. In equilibrium, profit taxation will, however, have distortionary effects through its impact on firm entry and exit, and the derived effects of this on firms' wage policies. As established above, see (2.13), income taxation has distortionary effects on workers' search effort and thus on the allocation of labor, see (2.14) and (2.16). It is evident from (2.22), (2.23) and (2.24) that income taxation therefore induce further equilibrium distortions of firms' wage and hiring policies.

### 2.2.3.2 Optimal wage and hiring policies

We restrict attention to a pure strategy equilibrium in which the solution to (2.23) and (2.24) are two functions  $w : [0, 1] \times [p_0, 1] \mapsto \mathbb{R}$  and  $v : [0, 1] \times [p_0, 1] \mapsto \mathbb{R}$  mapping worker-types and firm-types into wage offers and hiring intensities. In a similar, but simpler, model where search effort and hiring intensity are exogenously fixed at constant levels, Bontemps et al. (2000) show that, provided that the underlying productivity distribution  $\Gamma_0(\cdot)$  is continuous, the pure strategy equilibrium is in fact the only equilibrium. We conjecture that this result carries over to our slightly more complicated setup.

The optimal wage and hiring intensity policies,  $w(a|p)$  and  $v(a|p)$ , have a couple of useful and intuitive properties that are formally stated below.

**Proposition 1.**  *$w(a|p)$  is strictly increasing in  $p$  for all  $a \in [0, 1]$ .*

*Proof.* See Appendix A.1.6. □

**Proposition 2.**  *$v(a|p)$  is unique and is strictly increasing in  $p$  for all  $a \in [0, 1]$ .*

*Proof.* See Appendix A.1.7. □

**Proposition 3.**  *$\pi(p)$  is strictly increasing in  $p$ .*

*Proof.* See Appendix A.1.8. □

### 2.2.3.3 Firm entry and exit

Firm entry and exit is as in Melitz (2003). Firms enter the market by paying a sunk entry cost  $C$ . Our current model formulation assumes the presence of a third class of economic agent in the economy besides workers and firms, namely entrepreneurs. An entrepreneur is a risk neutral agent who therefore do not participate in the risk sharing arrangement in the large households. An entering firm is an entrepreneur who pays the sunk cost to develop a business idea. One might for example think of  $C$  as a money-metric disutility cost of an entrepreneurs foregone leisure. Indeed, we treat  $C$  as non-deductible in relation to profit taxation. Immediately upon entry, the firm observes its productivity  $p$ , and decides whether to hire workers and start producing, or exit. A firm's productivity  $p$  is drawn from the exogenous productivity distribution  $\Gamma(\cdot)$  with support  $[0, 1]$ . A firm with productivity  $p$  exits if the steady profit at that productivity is negative, i.e. if  $\pi(p) < 0$ . Proposition 3 established that  $\pi(p)$  is strictly increasing in  $p$ . Hence, there is a threshold productivity  $p_0$  such that firms with  $p < p_0$  exit the market. The threshold productivity  $p_0$  solves  $\pi(p_0) = 0$  and the endogenous distribution of productivities across operating firms,  $\Gamma_0(\cdot)$ , is given by

$$\Gamma_0(p) = \frac{\Gamma(p) - \Gamma(p_0)}{\bar{\Gamma}(p_0)}, \quad p \in [p_0, 1]. \quad (2.25)$$

Free entry ensures that, in equilibrium, the measure of active firms in the economy,  $M_0$ , adjust such that the expected steady state profit flow from entry is offset by the opportunity (flow) cost of entry, given by the annuitized entry cost  $\rho C$ . That is,

$$\int_{p_0}^1 \pi(p) d\Gamma(p) = \rho C. \quad (2.26)$$

Substituting (2.23) for  $\pi(p)$  in (2.26), and using that  $V(a) = M_0 \int_{p_0}^{\bar{p}} v(a|x) d\Gamma_0(x)$ , implies that

$$M_0 = \left[ \frac{\rho C}{1 - \tau} + \int_{p_0}^1 \int_0^1 d(v(a|p)) da d\Gamma(p) \right]^{-1} \left[ \int_{p_0}^1 \int_0^1 \frac{v(a|p) \tilde{\pi}(a|p) h(a)}{\int_{p_0}^1 v(a|x) d\Gamma_0(x)} da d\Gamma(p) \right]. \quad (2.27)$$

Equation (2.27) shows that profit taxation distorts firms' entry and exit decisions, and thus impact the measure of active firms. These distortion arise because entry costs are sunk, and therefore not deductible. Profit taxation effectively increases the cost of entry, or reduces the value of entry, thereby depressing the number of firms in the economy.

**Ownership structure of firms.** Firms are owned by the risk neutral entrepreneurs who simply consumes the net-of-tax profit flow, on average  $\rho C$ . As mentioned above, the entrepreneurs do not participate in the risk sharing arrangement of the risk averse workers. Indeed, the household budget constraint (2.4) does not include any net-of-tax profits. We stress that the entrepreneurs are within the jurisdiction of the tax authorities as the revenue from profit tax enters the government's budget (2.1).

An alternative model formulation assumes that the households are the ultimate owners of the firms, perhaps through a mutual fund. In such a formulation, net-of-tax profit flows accrue to households and enters the household budget constraint (2.4), and will have implications for workers' search behavior via income effects. While such a model formulation is perhaps more conventional in the macro literature, we note that the vast majority of individuals in our data in fact have negative capital income, see Figure 2.1.<sup>15</sup> Hence, from a purely empirical standpoint, it is not clear that capital income generates significant income effects in workers search behavior. Nonetheless, to explore these issues further we are currently working on a model formulation where household own the firms through mutual funds.

#### 2.2.3.4 Aggregation

Since firms optimal worker-type  $a$  conditional wage offers  $w(a|p)$  are strictly increasing in firm-type  $p$ , see Proposition 1, and firms meet potential employees at a rate proportional to hiring intensity  $v(a|p)$ , the wage offer distribution faced by workers is simply the hiring intensity weighted productivity distribution given by

$$F(w|a) \equiv F(w(a|p)|a) = \frac{\int_{p_0}^p v(a|x) d\Gamma_0(x)}{\int_{p_0}^1 v(a|x) d\Gamma_0(x)}. \quad (2.28)$$

---

<sup>15</sup>This result arise because the measured capital income include interest on mortgage debt. Taking out mortgage interest, we believe most individuals in our data have little or not capital income.

Let  $\theta(a)$  be the worker-type  $a$  specific labor market tightness, measured as the ratio of aggregate hiring intensity  $V(a) = M_0 \int_{p_0}^1 v(p) d\Gamma_0(p)$  to aggregate search effort  $S(a)$  in the worker-type  $a$  labor market. Aggregate worker-type  $a$  search effort  $S(a)$  is given by

$$S(a) = n^0(a)[\mu + s(b|a)] + [1 - n^0(a)] \int_b^{\bar{w}(a)} [\mu + s(w|a)] dG(w|a), \quad (2.29)$$

where  $s(w|a)$  is given by (2.13),  $n^0(a)$  is given by (2.14), and  $G(w|a)$  is given by (2.16), and where, by Proposition 1,  $\bar{w}(a) = w(a|1)$ , the optimal wage offer in the most productive firm with  $p = 1$ .

The model structure implies that the flow of contacts between type- $a$  workers and firms in the economy is  $\lambda S(a)$ . With a constant returns to scale Cobb-Douglas matching function with elasticity of matches with respect to unemployed search being  $\eta$ , we have that  $\lambda(a) = \theta(a)^{1-\eta}$ . This closes the model, and we can state the equilibrium conditions in the next subsection.

## 2.2.4 Equilibrium

Our equilibrium definition follows Mortensen (2005).

**Definition 1.** *A steady-state labor market equilibrium is composed of the following components:*

- *The optimal reservation strategy: An employed worker accepts any job offer that offers a wage strictly greater than the current wage. An unemployed type- $a$  worker accepts any job offer greater than or equal to  $\phi(a) = b$ , see (2.12).*
- *Workers' optimal search effort solves (2.13).*
- *Firms' worker-type specific optimal wage and hiring intensity policies  $w(a|p)$  and  $v(a|p)$  solve (2.24) and (2.23).*
- *The worker-type specific unemployment rate  $n^0(a)$  satisfies (2.14) and the distribution of wages across employed workers  $G(w|a)$  satisfies (2.16).*
- *The conditional gross wage offer distribution  $F(w|a)$  that satisfies (2.28).*
- *The threshold productivity for entry  $p_0$  solves  $\pi(p_0) = 0$  with  $\pi(p)$  given by (2.20).*
- *The measure of firms operating in the market  $M_0$  given by (2.27).*
- *The worker-type specific matching efficiency parameters  $\lambda(a)$  are given by  $\lambda(a) = \theta(a)^{1-\eta}$  where  $\theta(a) = S(a)/V(a)$ ,  $V(a) = M_0 \int_{p_0}^1 v(p) d\Gamma_0(p)$  and  $S(a)$  given by (2.29).*



Given the rich structure of the model, we have not been able to prove existence of the equilibrium. However, our solution algorithm finds a numerical solution under reasonable parameter values. The numerical model solution is detailed in Appendix A.4.

## 2.3 Data

Our empirical analysis uses a matched employer-employee panel covering the entire Danish population during 1990-2005. The data is constructed from administrative records. Estimation of the structural parameters of the model is carried out using a subset of this data pertaining to the period 1999-2003. Evaluation of a series of tax reforms implemented in 1994, 1998, and 2004 requires the use of the entire 1990-2005 data period.

### 2.3.1 Data sources

We build our matched employer-employee panel from three sources, (i) labor market spell data, including hourly wage information, (ii) information from IDA (Integreret Database for Arbejdsmarkedsforskning), a Danish register-based matched employer-employee database maintained by Statistics Denmark containing socioeconomic information on workers and some background information on firms, and (iii) information on firms' sales and purchases from firm-level value added tax (VAT) accounts administered by the Danish tax authorities.

**Labor market spell data.** The labor market spell data contains individual job and non-employment spells. Information on job spells is available for the period 1985-2013 for all legal Danish residents aged 15-74, and is obtained by combining a number of administrative registers.<sup>16</sup> We restrict attention to the period 1990-2005. A job spell is defined as a continuous period of primary employment at a given firm, with duration measured in days.<sup>17</sup> Job spells, and therefore labor market transitions, are recorded at the firm-level. A firm may consist of multiple workplaces, but the labor market spell data is constructed such that continuous employment at different workplaces within a firm is considered as a single job spell. Nonemployment spells are periods where no job spells are recorded. We are not able to distinguish between nonemployed who actively search for a job and nonparticipants.

---

<sup>16</sup>Henning Bunzel at Aarhus University has been instrumental in developing the job spell data described here.

<sup>17</sup>Primary attachment is evaluated calendar month by calendar month. For each individual in each month, the primary employer is defined as the firm at which the individual works the highest number of hours in the current and next two calendar months.

The unit of observation in the labor market spell data is a person-spell-year. Hence, a spell that starts on Jan. 5th, 2000 and ends on March 12th, 2002 is represented by three observations relating to the years 2000, 2001 and 2002. The job spell data includes worker and firm identifiers, start- and end-dates of the job, the annual earnings pertaining to the job, as well as an estimate of the annual number of hours worked in the job.<sup>18</sup> Hence, we can compute (an estimate of) the average annual wage rate in each job. In cases where a worker transitions from one firm to another with up to 14 days of nonemployment between the two employment spells, we overwrite the nonemployment spell and thus record a job-to-job transition. In cases where a worker ends an employment spell at a firm, but transitions back to the same firm with at most 12 weeks of non-employment between the two employment spells, we overwrite the nonemployment spell and instead record a single employment spell.

**IDA data.** IDA consists of several sub-panels. We use the sub-panels IDA-P and IDA-S from the period 1990-2005. IDA-P contains annual background information from public registers on all individuals aged 15-74 residing legally in Denmark on the 31st of December. IDA-P information includes highest completed education including date of completion, and information on any ongoing education subsequent to the highest completed education. We use date of completion of the highest completed education is used to define labor market entry. The unit of observation in IDA-P is a person-year. IDA-S contains background information from public records on all physical workplaces in Denmark. We retain a public sector indicator from IDA-S, defining a firm’s public sector status to be the public sector status of its largest workplace. The unit of observation in the (aggregated) IDA-S panel is a firm-year.

**VAT data.** Data on firms’ (not workplaces) sales and purchases are obtained from the panels MOMS and MOMM, constructed from firm’s value added tax (VAT) accounts maintained by the Danish tax authorities. Danish firms settle VAT accounts monthly, quarterly or annually depending on the size of their revenue. MOMS covers the period 1995-2000 and contains annual sales and purchases. MOMM is a monthly panel starting in January, 2001. We use MOMM data up until December, 2005, aggregating the monthly information to an annual frequency.<sup>19</sup> Putting MOMS and MOMM together, we obtain an

---

<sup>18</sup>Annual hours are estimated using information on mandatory pension contributions. Lund and Vejlin (2015) develop and implement a procedure for computing annual hours in a job in the IDA data for the period 1980-2007, primarily using information on mandatory pension contributions. We adapt this procedure for the spell data with some minor simplifications.

<sup>19</sup>For firms that settle VAT accounts on quarterly or annual frequencies, the monthly information in MOMM are imputed by Statistics Denmark. As all firms settle VAT accounts at least annually, the annual aggregate data is not affected by the imputation.

annual panel of firms' sales and purchases for the period 1995-2005. We compute annual firm-level value added as annual sale less annual purchases. The unit of observation in the VAT panel data is a firm-year.

**Merging the data sources.** Merging the data sources is relatively straightforward using the available person and firm identifiers, and Appendix A.2 provide a detailed description of the merging procedure. Here we simply note that the resulting 1990-2005 matched employer-employee panel contains 84,474,045 spell observations on 4,447,401 individuals, 428,448 firms, 57,568,393 job spells and 26,905,652 non-employment spells prior to selection of the analysis data.

### 2.3.2 Selecting the analysis data

We start by discarding all observations on individuals never observed with either age or education information.<sup>20</sup> We also discard individuals with implausible education information, defined as age minus years of schooling being less than 5. We then define labor market entry to occur at the observed date of graduation from highest completed education (as observed in the window of observation, 1990-2005), or at January, 1st of the year an individual turns 19, whichever occur at the latest date. All pre-entry observations are discarded. We define labor market exit to occur at the last year an individual is observed residing legally in Denmark (i.e. present in IDA-P) or at December, 31st of the year an individual turns 55, whichever occur first. All post-exit observations are discarded. We right censor observations at age 55 to avoid labor market behavior driven by retirement considerations.

A series tax reforms were implemented in Denmark on January 1st in 1990, 1994, 1999 and 2004. As we document further below, between these reform dates, the tax system is relatively stable, and we therefore extract four short matched employer-employee panels pertaining to the periods 1990-1993, 1994-1998, 1999-2003, and 2004-2005. We employ a rudimentary, but consistent, set of selection criteria to each of the four panels. These criteria are imposed independently for each panel. Hence, an individual or firm discarded from, say, the 1999-2003 panel due to violation of the selection criteria to be outlined below, will be included in the other three analysis datasets if the same selection criteria are not violated during in the 1990-1993, 1994-1998 or 2004-2005 panels.

Effectively, we select individuals who primarily work full time in the private sector (when they work). Specifically, for each of the four panels covering 1990-1993, 1994-1998, 1999-2003, and 2004-2005, we discard workers who, (i) in any year worked less than 25

---

<sup>20</sup>The primary cause of missing education data is foreign educational credentials. Hence, we are likely to discard predominantly immigrants and refugees in this step.

Table 2.1: November 28th cross-sections in the 1999-2003 estimation panel

Year	# Persons	# Firms	Nonemploy- ment rate
1999	932,868	75,310	28.9%
2000	895,223	71,710	28.8%
2001	861,383	71,890	28.6%
2002	831,835	70,423	28.7%
2003	813,555	70,902	30.3%

hours a week on average, when working, (ii) at any point in time work in the public sector, or (iii) leave the panel prematurely (likely caused by periods of living abroad, or death). Finally, for the estimation period 1999-2003, we trend value added and wages to 2003 prices using the internal wage deflator.

The selection process leaves us with four short matched employer-employee panels for the periods 1990-1993, 1994-1998, 1999-2003, and 2004-2005. Table 2.1 provides basic summary statistics for the estimation period 1999-2003. We note that the number of workers and firms, as well as the non-employment rates is relatively stable. Table 2.2 report some further basic descriptive statistics on the estimation dataset. On average, a worker is observed for slightly more than 4 years, during which she works at 1.84 firms. The descriptive statistics on age and education is in line with aggregate statistics for the Danish labor force.

## 2.4 The empirical tax function

The income tax function is a key component in our analysis. This section provides a brief overview of the Danish income tax system, discusses the 1994, 1999, and 2004 tax reforms, and describe how we estimate marginal and average tax functions using the micro-level analysis panels described above.

### 2.4.1 The Danish income tax system 1990-2005

**An overview.** Income taxes in Denmark are divided into regional and national taxes. Regional taxes include municipality tax, county tax, and a minuscule church tax paid only by members of the Church of Denmark. Regional tax rates differ by municipality and county, although the variation in tax rates is small. National taxes have a progressive structure with three brackets (bottom, middle and top) with tax rates cumulated across brackets. Regional and national taxes differ not only in tax rates, but also in terms of the

Table 2.2: Summery Statistics for the 1999-2003 estimation panel

	Mean	Std. Dev.	Percent
Number of years in sample	4.34	1.29	
Number of workers per firm	8.53	93.16	
Number of firms per worker	1.84	1.24	
Age	41.30	8.93	
Highest completed education			
Primary			36.75
High school			6.06
Vocational education			44.38
Short tertiary education			3.20
Medium tertiary education			5.86
Bachelor			0.57
Master degree or longer			3.17

Note: Numbers of years in the sample is computed as the number of years (maximum five) a given worker is observed in the sample. Number of workers per firm is the number of workers a given firm employs at a given November cross section. Number of firms per worker is the number of firms a given worker is observed with during the panel.

Table 2.3: Income concepts in the Danish income tax system

Income concept	Main items included
Labor income ( $LI$ )	Salary, wages, honoraria, fees, bonuses, fringe benefits, business earnings
Personal income ( $PI$ )	$LI$ + transfers, grants, awards, gifts, received alimony – labor market contribution, certain pension contributions
Capital income ( $CI$ )	Interest income, rental income, business capital income – interest on debt (mortgages, bank loans, credit card debt, student loans)
Deductions ( $D$ )	Commuting, union fees, UI contributions, other work expenditures, charity, paid alimony

Note: Reproduction of Table 1 in Kleven and Schultz (2014b).

base on which they are levied. Furthermore, during the period we consider, various social security and labor market contributions (SSC and LMC), as well as earned income tax credits (EITCs) have been introduced (and in some cases abolished again).<sup>21</sup> Finally, a marginal tax ceiling is in place. If the sum of all tax rates, excluding the church tax, and any social and labor market contributions, exceeds this ceiling, the top tax rate is adjusted to equal the ceiling. The Danish tax code is based on individual filing for married couples, but some components of deductions or allowances can be transferred across spouses. We take the spousal income into account when we simulate marginal tax rates.

The tax base of regional and national income taxes consists of a combination of income concepts, including personal income, capital income, and deductions. Table 2.3 details the main items in each of the taxable income components relevant for the taxation of labor income during 1990-2005, and Table 2.4 provides descriptive statistics on the distribution of each of these components in the 1999-2003 estimation data.<sup>22</sup> We note that, on average, labor income dwarfs the other income types, as well as the deductions, but also that there are considerable variation in each of the taxable income components, highlighting the benefit of having comprehensive micro data at our disposal. Finally, capital income is negative for the majority of Danish taxpayers as a result of interest payments on mortgages and loans.

<sup>21</sup>Social security contributions are compulsory payments that confer entitlement to future social benefit, including unemployment benefits. For example, a labor market contribution introduced in 1994, was at the onset effectively a social security contribution. Earmarking of the labor market contributions ceased in 2008, after which it is best described as a regular (flat) income tax.

<sup>22</sup>Since 1987, the taxation of wealth including income from stock holdings is completely separate from the taxation of labor income. Consistent with the theoretical model our empirical analysis ignores the role of wealth in shaping labor market behavior. See Jakobsen, Jakobsen, Kleven, and Zucman (2018) for a comprehensive analysis of wealth taxation in Denmark.

Table 2.4: The summary statistics on the distribution of 1999-2003 taxable income components

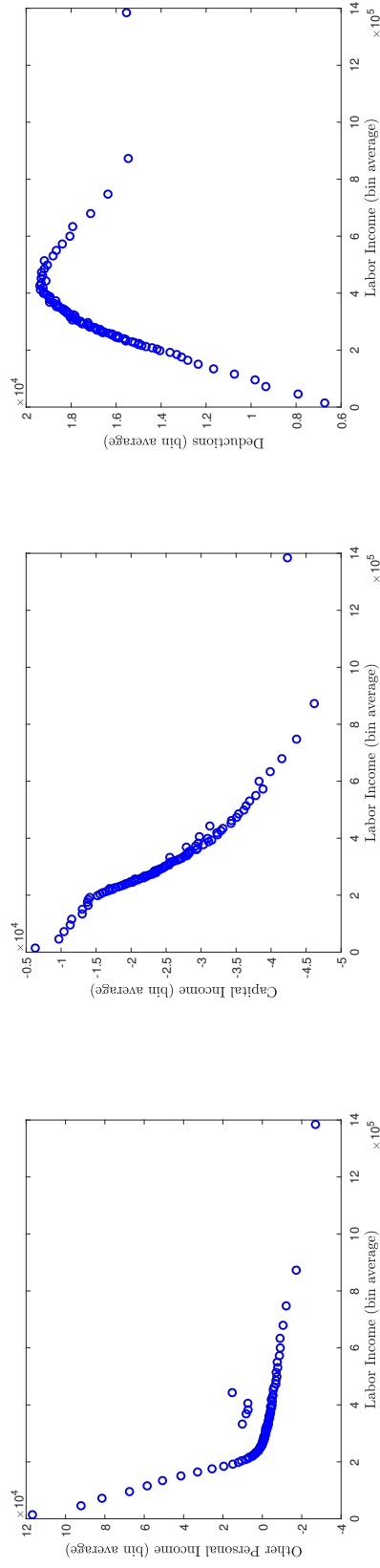
	Average	Std. dev.	P25	P50	P75
Labor income	333,763	184,862	245,640	301,755	381,665
Personal income net of labor income	4,485	590,762	-3,752	1	2,001
Capital income	-25,158	99,442	-41,549	-22,281	-4,288
Deductions	16,691	13,579	9,489	13,461	19,719

Note: All amounts denominated in 2003 Danish Kroner. P25, P50 and P75 refers to the 25th, 50th and 75th percentiles in the distribution of each of the variables (pooled across the years 1999-2003)

To get a sense of the correlation structure between the different income concepts, Figure 2.1 plots the average personal income net of labor income, capital income and deductions against the percentiles in the distribution of labor income for the 1999-2003 panel used in our estimation. From panel A, we see that, naturally, personal income net of labor income is, on average, higher for individual with low labor income, reflecting primarily transfers and benefits related to periods of nonemployment. However, once labor income reaches about 250,000 Danish Kroner—around the 25th percentile in the labor income distribution—the personal income net of labor income component is, on average, close to zero, and remains small as we move through the labor income distribution. In fact, personal income net of labor income becomes, on average, negative for high labor incomes, most likely reflecting certain pension contributions. Panel B in Figure 2.1 confirms that capital income is, on average, negative for all labor incomes, and that capital income is negatively related to the labor income component. This relationship reflects that, for most individuals, capital income consists almost exclusively of interest on mortgage debts, and that individuals with higher labor income tend to have larger mortgages. Still, the ratio of capital income to labor income remains small throughout the distribution of labor incomes. Finally, Panel C in 2.1 shows that deductions are a hump-shaped function of labor income, and that the amounts deducted are small relative to labor income throughout the distribution of labor incomes.

Overall, the descriptive evidence in Table 2.4 and in Figure 2.1 clearly indicate that labor income is the primary source of income for almost all individuals in our data, and that, for most individuals, capital income and deductions will have not greatly affect their available permanent income. We will account for the presence of non-labor income components in the estimation of the tax function detailed below, but our model ignores any consumption enjoyed out of non-labor income. The evidence in Table 2.4 and in Figure 2.1 suggest that the bias induced by this omission is likely to be small.

Figure 2.1: Average personal income, capital income, and deductions by labor income percentiles



Panel A: Personal income net of labor income

Panel B: Capital income

Panel C: Deductions

Note: Each of the figures plot averages computed within the percentiles of the distribution of labor income in the pooled 1999-2003 cross sections against the percentiles of the labor income distribution. All amounts denominated in 2003 Danish Kroner.



**Income tax reforms.** There were three major reforms of the income tax system in Denmark during 1990-2005, namely in 1994, 1999, and 2004.<sup>23</sup> Although some of the reforms were part of a larger fiscal reform packages, all three reforms were intended to increase the return to work. The reforms implemented changes to the tax bases of regional and national taxes, as well as the tax rates, and as we shall see, induced significant shifts in the marginal income tax function faced by individuals. Table 2.5 describes the tax bases and rates in place during the periods 1990-1993, 1994-1998, 1999-2003, and 2004-2005. We note that the three reforms were characterized by (i) changes in tax bases in all national tax brackets, (ii) differential changes in progressive tax rates across brackets, (iii) abolition of social security contributions and introduction of labor market contributions and EITCs, and (iv) rising regional tax rates and declining marginal tax ceilings. We show further below that the the income levels defining bottom, middle and top tax brackets also change during the period (over and above indexation to inflation).

**The 1994 tax reform.** The 1994 reform was legislated in 1993 with implementation starting January 1st, 1994, but not completed until 1998. Following recommendation from a commission of appointed experts published in 1992, the purpose of the 1994 tax reform was to improve incentives to work, and to reduce opportunities for tax avoidance. While the 1994 reform largely followed the 1992 expert recommendations, it was considered key element in the the newly formed centre-left government attempts to stimulate the Danish economy. Besides the changes to the income tax bases and rates described in Table 2.5, the reform made social pensions and cash benefits tax liable to facilitate comparison to earned income. Finally, the 1994 reform involved a reduction in property taxes, and an increase in the use of levies and fees in environmental regulation.

**The 1999 tax reform.** The 1999 income tax reform was an integral part of wider set of economic reforms known as the “Pentecost reforms” (in Danish: “Pinsepakken”) due to the date it was proposed. The reform was legislated in June 1998, with the income tax reforms implemented January 1st, 1999. The reformation of the tax system was motivated by a political desire within the centre-left government to lower tax rates on low incomes, and to increase the marginal tax ceiling. Besides the changes to the income tax bases and rates described in Table 2.5, the “Pentecost reforms” made changes to property taxation, and further promoted the use of levies and fees in environmental regulation as a source of public revenue.

---

<sup>23</sup>The description of the reforms to come is based on material available on the website of the Danish Ministry of Taxation, see <http://www.skm.dk/aktuelt/temaer/skattereformer-og-skatteomlaegninger-siden-1987> (in Danish).

Table 2.5: Danish income tax bases and rates 1990-2005

Tax type	1993			1998			2001			2005		
	Base	Rate (%)		Base	Rate (%)		Base	Rate (%)		Base	Rate (%)	
Regional taxes	PI + CI - D	30.2 (avg)		PI + CI - D	32.4 (avg)		PI + CI - D	33.2 (avg)		PI + CI - D	33.3 (avg)	
National taxes												
Bottom bracket	PI + CI - D	22.0		PI + CI - D	8.0		PI + [CI > 0] - D	6.25		PI + [CI > 0]	5.5	
Middle bracket	PI + [CI > 0]	6.0		PI + CI	6.0		PI + [CI > 0]	6.0		PI + [CI > 0]	6.0	
Top bracket	PI	12.0		PI + [CI > 21k]	15.0		PI + [CI > 0]	15.0		PI + [CI > 0]	15.0	
SSC	-	-		-	-		-	-		-	-	
LMC	-	-		LI	8.0		LI	8.0		LI	8.0	
EITC	-	-		-	-		-	-		LI	2.5	
Marginal tax ceiling		68.0			58.0			59.0			59.0	

Note: Tax rates are cumulative. For example, an individual in the middle bracket in 1992 faces a marginal tax rate on labor income of  $33.20\% + 6.25\% + 6.00\% = 45.45\%$ . The 1994 tax reform was gradually implemented over the period 1994-1998. The tax bases and rates reported for 1994-1998 are those in place in 1998 when the reform was fully implemented. SSC is an acronym for Social security contributions. LMC is an acronym for Labor market contributions. EITC is an acronym for Earned income tax credits. For the tax base acronyms, see Table 2.3.

**The 2004 tax reform.** The 2004 income tax reform was designed by a centre-right government that had come to power in November 2001. The reform was motivated by desire to incentivize labor market participation and work by lowering the tax on labor income. Compared to the 1994 and 1999 reforms, the changes to the marginal tax rates on labor income was not offset by increases in other fees or levies due to the centre-right government having committed to not increase any taxes, levies or fees upon taking office. Besides the changes to the income tax bases and rates described in Table 2.5, the 2004 tax reform made small changes to vehicle excise duties, and committed public funds to social care and health care, particularly focused on the elderly.

## 2.4.2 Estimating the marginal tax functions

**Simulating individual marginal labor income tax rates.** Table 2.5 describes how an individual’s marginal tax rate depend on several factors, including the distribution of income across the income concepts listed in Table 2.3, region of residence, marital status, and, if cohabiting or married, partner income. Fortunately, our data contains all the information necessary to simulate the marginal tax rates faced by individuals in our data. The simulations are carried out using a simulator of the Danish tax system constructed by le Maire and Schjerring (2013) and Kleven and Schultz (2014b).<sup>24</sup>

The tax simulator allow us to compute the total tax liability for each individual, for each year that we consider (i.e. 1990-2005). For the purpose of this paper, where we focus on labor income taxation, it is useful to represent individual  $i$ ’s tax liability in year  $t$  as  $T_t(LI_{it}, \mathbf{z}_{it})$ , where  $T_t(\cdot)$  is the tax function in year  $t$  according to the tax simulator,  $LI_{it}$  is individual labor income, and  $\mathbf{z}_{it}$  is a vector of other income concepts relevant for the tax liabilities in year  $t$ , see Table 2.3. Following Kleven and Schultz (2014b) we compute individual marginal tax rates as

$$T'_t(LI_{it}, \mathbf{z}_{it}) = \frac{T_t(LI_{it} + 100, \mathbf{z}_{it}) - T_t(LI_{it}, \mathbf{z}_{it})}{100}.$$

That is, we add 100 Danish Kroner (approximately 15 US Dollars in 2003) to annual labor income  $LI$ , and record the share of this marginal increase that is appropriated by taxes. We stress that marginal tax rates are computed individual by individual, year by year, and reflects individual circumstances such as the composition of income across the income concepts Table 2.3, region of residence, and relevant household characteristics. Hence,

---

<sup>24</sup>We are grateful to Henrik Kleven and Esben Schultz for making the tax simulator available. The tax simulator consists of a set of SAS-programs which we downloaded from the website of the *American Economic Journal: Applied Economics*. We collected data on regional taxes in Denmark for the period 1985-2005 to use as an input in the tax simulations. This data is available in online supplementary material to this paper.

the simulated marginal tax rates represent the actual marginal tax rate faced by each individual in our data.

**Estimating the marginal labor income tax function.** We use the annual individual marginal tax rates as data points for estimation of a marginal labor income tax function for each tax regime, i.e. for each of the periods 1990-1993, 1994-1998, 1999-2003, and 2004-2005. As mentioned above, the 1994 tax reform was implemented gradually over the period 1994-1998. For that particular tax regime, we therefore estimate a marginal labor income tax function for each year. For the remaining three tax regimes, including our estimation period 1999-2003, we pool the annual marginal tax rates and estimate a single marginal labor income tax function. To facilitate the pooling, we deflate labor income to 2003 levels using the internal deflator from spell data wages described above.

Our structural model of worker search requires a marginal taxes as a function of hourly wages, not annual labor income. We transform the observed labor income into hourly wages by dividing by a measure of annual full-time working hours, as constructed by Lund and Vejlin (2015).<sup>25</sup>

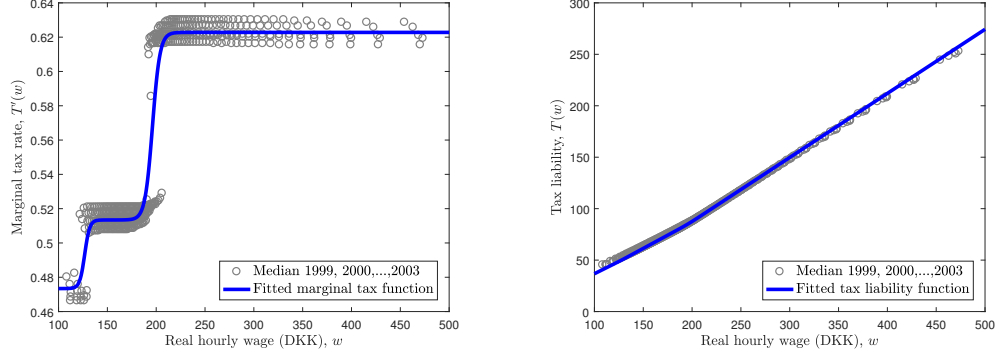
As individual tax liabilities  $T_t(LI_{it}, \mathbf{z}_{it})$  are not solely functions of labor income  $LI$ , but also depend on other income components, e.g. capital income, and household characteristics, the mapping from hourly wages to marginal tax rates is not one-to-one. To proceed with estimation of the marginal labor income tax functions, we split hourly wages into 100 bins representing the percentiles in the empirical distribution of hourly wages for a given year in a given tax regime. We next compute the median marginal tax rates for each hourly wage bin in each year, pool the annual median marginal tax rates and fit a flexible parametric function (of hourly wages) to these median median marginal tax rates.<sup>26</sup> Figure 2.2 shows the resulting marginal tax function (Panel A) and tax function (Panel B) for the estimation period 1999-2003. The marginal tax function in Panel A is superimposed on the annual within-wage-percentile median tax rate, and the average tax function in Panel B is obtained by integrating the marginal tax function. Similar plots of the marginal tax functions and tax functions for the periods 1990-1993, 1994-1998, and 2004-2005 are presented and discussed further below in relation to our tax reform evaluation exercise.

---

<sup>25</sup>According to this measure, annual hours range between 1,665 and 1,695 for the estimation period 1999-2003.

<sup>26</sup>Appendix A.3 provides further details on this procedure.

Figure 2.2: Marginal tax function and tax function for Denmark 1999-2003



Panel A: Marginal tax function  $T'(\cdot)$

Panel B: Tax function  $T(\cdot)$

Note: Panel A plots the marginal tax function  $T'(w)$  and Panel B plots the average tax function  $T(w)$ . The gray circles in Panel A represent the annual within-wage-percentile median tax rates as described in the text. The gray circles in Panel B represent the annual within-wage-percentile median hourly tax liabilities.

## 2.5 Model estimation

Armed with the empirical tax function, we proceed to estimation of the model's structural parameters by way of Indirect Inference, see Gourieroux, Monfort, and Renault (1993). The estimation procedure finds the set of structural parameters which best fit a set of chosen auxiliary statistics. The model is identified if there is one and only one set of structural parameters which generates the auxiliary statistics of the estimation. This section describes how we parameterize the model, discusses issues related to identification of the model parameters, reports and discusses the estimated parameters as well as the model's ability to fit the chosen auxiliary statistics.

### 2.5.1 Parameterization

The presentation of the model in section 2.2 left some key objects, such as the exogenous heterogeneity distributions, match output and workers' felicity function, not fully parametrically specified. For the purpose of estimation, specific parametric forms are required.

Match output  $y(a, p)$  is specified as a flexible second order polynomial. Indeed,

$$y(a, p) = \varrho_0 + \varrho_1 a + \varrho_2 p + \varrho_3 a^2 + \varrho_4 p^2 + \varrho_5 ap. \quad (2.30)$$

where  $\boldsymbol{\varrho} = (\varrho_0, \dots, \varrho_5)$  are unknown parameters. Consistent with the theoretical model,

estimation is carried out under the restriction  $y_a(a, p) > 0$  and  $y_p(a, p) > 0$ .<sup>27</sup> The match output function (2.30) is sufficiently flexible to warrant a normalization of the exogenous worker and firm heterogeneity distributions to be uniform. That is,  $H(\cdot) = \mathcal{U}[0, 1]$  and  $\Gamma(\cdot) = \mathcal{U}[0, 1]$ .

Worker felicity (2.2) has two components, namely utility from consumption, represented by  $u(\cdot)$ , and disutility from job search effort, represented by  $\zeta(\cdot)$ . The utility function  $u(\cdot)$  is taken to be logarithmic, i.e.

$$u(c) = \log(c), \quad (2.31)$$

which correspondes to a moderate elasticity of intertemporal substitution. The search effort disutility component is specified as an isoelastic function  $\zeta(s)$ ,

$$\zeta(s - \underline{s}) = \frac{\zeta_0(s - \underline{s})^{1+1/\zeta_1}}{1 + 1/\zeta_1} \quad (2.32)$$

which accounts for free search  $\underline{s}$ , and where  $\zeta_0 > 0$  and  $\zeta_1 > 0$  are parameters to be estimated. Recall that the curvature of  $\zeta(\cdot)$ , determined by the parameter  $\zeta_1$ , is a key determinant of the substitution effect in job search effort.

A brief comment on the restrictive logarithmic specification of the utility function is warranted here. The theoretical analysis made it clear that the curvature of the utility function is a key driver of workers' responses to changes in the tax regimes. We do, however, currently not attempt to estimate this curvature. We are working on adding survey data on consumption expenditures to our data which we conjecture will allow us to identify a more general utility function. We are also currently working on establishing whether identification of the curvature in the utility function can be obtained from the already available mobility and wage data. For now, however, the utility function is restricted to be logarithmic, as per (2.31).

The hiring intensity cost function  $\chi(\cdot)$  is specified in an analogous fashion, i.e.

$$\chi(v) = \frac{\chi_0 v^{1+1/\chi_1}}{1 + 1/\chi_1}, \quad (2.33)$$

where  $\chi_0 > 0$  and  $\chi_1 > 0$  are parameters to be estimated. The isoelastic search effort disutility and hiring cost functions are standard in the literature, see e.g. Christensen et al. (2005), Kreiner et al. (2015), and Bagger and Lentz (2015).

For empirical realism we enrich the model to incorporate ability-dependent job de-

---

<sup>27</sup>In addition, we are currently imposing the restriction  $y(0, 0) = \varrho_0 = b$  in the estimation. Since job search off- and on-the-job are equally efficient, the latter restriction ensures that a match between the least able worker ( $a = 0$ ) and the least productive firm ( $p = 0$ ) is viable, and that all workers accept all jobs.

struction rates. Specifically, we let

$$\delta(a) = \exp(\delta_0 + \delta_1 a), \quad (2.34)$$

and estimate the parameters  $\delta_0$  and  $\delta_1$ . This completes the parameterization of the model.

The fully parameterized model features the following structural parameters  $(\rho, b, \underline{\varrho}, \zeta_0, \zeta_1, \underline{s}, \chi_0, \chi_1, \delta_0, \delta_1, \mu, C, \eta)$ . We remind the reader that  $\rho$  is the discount rate,  $b$  is the unemployment flow income,  $\mu$  is the reallocation rate,  $C$  is the entry cost, and  $\eta$  is the elasticity of the aggregate matching function with respect to job search effort. We fixed a few of the structural parameters prior to estimation. Specifically, we set  $\rho = 0.05$  per year, and we also fix the unemployment flow income  $b = \varrho_0 = 107$  Danish Kroner (DKK) per hour. In addition,  $\zeta_0$  and  $\chi_0$  are not separately identified as they both scale the baseline job finding rate  $\lambda$ . We therefore impose the normalization  $\chi_0 = 1$ ,<sup>28</sup> which leaves fifteen structural parameters to be estimated. Collect these in the vector  $\boldsymbol{\omega} = (\varrho_1, \dots, \varrho_5, \zeta_0, \zeta_1, \underline{s}, \chi_0, \chi_1, \delta_0, \delta_1, \mu, C, \eta)$ . Let  $\boldsymbol{\omega}_0$  be the true value of  $\boldsymbol{\omega}$ . The indirect inference estimator is given by

$$\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega}} [\mathbf{a}(\boldsymbol{\omega}_0) - \mathbf{a}^S(\boldsymbol{\omega})]' \boldsymbol{\Omega} [\mathbf{a}(\boldsymbol{\omega}_0) - \mathbf{a}^S(\boldsymbol{\omega})], \quad (2.35)$$

where  $\boldsymbol{\Omega}$  is a symmetric and positive definite weighting matrix, and  $\mathbf{a}(\boldsymbol{\omega}_0)$  is a vector of auxiliary statistics computed using the data described in section 2.3, i.e. at the unknown true value of the structural parameter vector  $\boldsymbol{\omega}_0$ . The vector  $\mathbf{a}^S(\boldsymbol{\omega}) = S^{-1} \sum_{s=1}^S \mathbf{a}_s(\boldsymbol{\omega})$  contains the same auxiliary statistics, but computed using  $N^s$  labor market histories simulated from the structural model for some value  $\boldsymbol{\omega}$  of the structural parameter value. Effectively, the Indirect Inference estimator finds the structural parameter vector that minimizes the distance between the real data and data simulated from the structural model, as measured by the included auxiliary statistics, and the weight matrix  $\boldsymbol{\Omega}$ .<sup>29</sup> Appendix A.4 contains a detailed description of the numerical computation of the model equilibrium, as well as a description of the simulation procedure we apply.

Implementation of (2.35) requires first of all a set of auxiliary statistics, and we discuss the ones we include below. In addition, however, implementation requires us to set  $S$ , the number of simulation repetitions,  $N^s$ , the number workers (i.e. the number of simulated labor market histories per simulation repetition  $s$ ), as well as the weight-matrix  $\boldsymbol{\Omega}$  for

<sup>28</sup>The same argument applies to any multiplicative constants in the aggregate matching function, which is also implicitly normalized to unity.

<sup>29</sup>Under mild regularity conditions, see Gourieroux et al. (1993), we have  $\sqrt{N}(\hat{\boldsymbol{\omega}}^S - \boldsymbol{\omega}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (1 + S^{-1})\mathbf{V}_0)$ , where  $\mathbf{V}_0 = [\mathbf{J}_0' \boldsymbol{\Sigma}_0 \mathbf{J}_0]^{-1} \mathbf{J}_0' \boldsymbol{\Sigma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Sigma}_0 \mathbf{J}_0 [\mathbf{J}_0' \boldsymbol{\Sigma}_0 \mathbf{J}_0]^{-1}$ ,  $\mathbf{J}_0 = \partial \mathbf{a}(\boldsymbol{\omega}) / \partial \boldsymbol{\omega}$  at  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ ,  $\boldsymbol{\Sigma}_0$  is the variance-covariance matrix of  $\mathbf{a}(\boldsymbol{\omega}_0)$ . We report standard errors based on the asymptotic variance-covariance matrix  $N^{-1}(1 + S^{-1})\hat{\mathbf{V}}$ , where  $\hat{\mathbf{V}}$  is  $\mathbf{V}_0$  with true values replaced by estimates.  $\boldsymbol{\Sigma}_0$  is estimated by block-bootstrapping  $\mathbf{a}(\hat{\boldsymbol{\omega}})$ , with workers' labor market histories constituting the blocks. We take  $N$  to be the number of workers in our estimation data.

computing the distance between empirical and simulated auxiliary statistics, see (2.35). In our implementation, we take  $S = 1$ ,  $N^s = 100,000$ , and take  $\mathbf{\Omega}$  to be a diagonal matrix. With a few exceptions reported further below, the diagonal entries in  $\mathbf{\Omega}$  are set to 1.<sup>30</sup>

### 2.5.2 The government budget

In steady state, a type- $a$  worker's contribution to the government budget is

$$B(a) = n^0(a)T(b) + [1 - n^0(a)] \int_0^1 T(w(a|p))dG(w(a|p)|a) - n^0(a)b - \bar{b}, \quad (2.36)$$

where  $\bar{b}$  is a lump-sum transfer, see (2.1). The government budget is therefore  $B = \int_0^1 B(a)dH(a)$ .

Ideally, estimation of the model's structural parameters should be conducted under a balanced budget restriction, i.e.  $B = 0$ , achieved through the lump-sum transfers  $\bar{b}$ . However, these transfers are not straightforward to implement in our model because a) we would need to take a stance on who receives the lump-sum transfers, which in the absence of a government welfare function is somewhat arbitrary, and, more importantly, b) because recipients of government transfers will adjust their search effort as the income effects kick in. This, in turn, sets in motion a sequence equilibrium adjustments, which feeds back into the government budget. We are currently augmenting our equilibrium computation algorithm to allow for lump-sum transfers that balance the government budget. In doing so, we distribute any government surplus or deficit equally across workers, independently of ability. The current estimates, however, does not allow for lump-sum transfers, and hence, does not balance the government budget. That is, the current estimates impose  $\bar{b} = 0$ . To nonetheless facilitate interpretation, we always report changes in the government budget associated with various actual or counterfactual income tax reforms.

### 2.5.3 Auxiliary statistics

The estimation panel contains primarily information on wages and labor market transitions, and both these features of the data are needed to identify the model. In addition, the data has information on value added, which we also make use of. The description of the auxiliary statistics provided here is nontechnical and focuses on the intuition underlying the identification arguments for each of the structural parameters.

---

<sup>30</sup>Asymptotic efficiency of the Indirect Inference estimator requires  $\mathbf{\Omega} = \hat{\mathbf{\Sigma}}^{-1}$ , i.e. using the inverse variance-covariance matrix of the vector of auxiliary statistics as weight-matrix. Altonji and Segal (1996) illustrates how an asymptotically efficient GMM estimator may be marred by small sample biases in applications.



Given our focus on a labor market with two-sided heterogeneity, it is useful to represent wages in a way that captures this feature of the data.<sup>31</sup> For this reason, let  $i$  index individuals,  $j$  index firms, and let  $n$  index annual November 28th cross sections. In addition, let  $J(i, n) = j$  if worker  $i$  in cross section  $n$  is employed by firm  $j$ , and consider the following auxiliary two-way fixed effect log-wage regression based on Abowd, Kramarz, and Margolis (1999),

$$\ln w_{in} = \phi_i + \psi_{J(i,n)} + \varepsilon_{in}, \quad (2.37)$$

where, here,  $\phi_i$  is a person effect,  $\psi_j$  is a firm effect, and  $\varepsilon_{in}$  represent residual log-wage variation. As indicated by the notation, we estimate this auxiliary model of wages using repeated annual November 28th cross sections. Estimation of the worker and firm fixed effects is done by ordinary least squares imposing the assumption  $E[\varepsilon_{in}|i, J(i, n)] = 0$ , and we ensure that the design matrix of regressors, containing only worker and firm dummies, has full rank by estimating (2.37) only on the largest set of connected workers and firms, see Abowd, Creecy, and Kramarz (2002) for details.<sup>32</sup>

The auxiliary wage regression (2.37) allows a decomposition of log wages into a worker-component  $\hat{\phi}_i$ , a firm-component  $\hat{\psi}_j$ , and a residual component  $\hat{\varepsilon}_{in} \equiv \ln w_{in} - \hat{\phi}_i - \hat{\psi}_{J(i,n)}$ . Indeed,  $\ln w_{in} = \hat{\phi}_i + \hat{\psi}_{J(i,n)} + \hat{\varepsilon}_{in}$ . These components form the basis of many of the auxiliary wage statistics we include in the estimation. We stress that, even though (2.37) is not derived from the structural model and therefore likely is misspecified, the resulting decomposition of wages is useful because it yields measurements related to the underlying heterogeneity worker- and firm types,  $a$  and  $p$ .<sup>33</sup> We normalize the grand mean of the estimated firm effects to zero.

**Identification of the match-level output function  $y(a, p)$ .** The model-predicted wage offers  $w(a|p)$  are complicated equilibrium objects that depends on all parameters of the model, including the match-level output function  $y(a, p)$ . However, holding search effort and hiring intensity fixed—and we argue below that search and hiring behavior is identified from data sources unrelated to wages—the distribution of wage offers is largely determined by the characteristics of the output function  $y(a, p)$ . Hence, statistics describing the distribution of wages contains identifying information regarding the parameters  $\boldsymbol{\varrho}$

<sup>31</sup>Unlike wages, value added is available only at the firm-level, not at the match-level. Given the focus on two-sided heterogeneity, this limits the identifying power from value added data.

<sup>32</sup>In our data, as well as in our model simulations, almost all workers and firms are connected.

<sup>33</sup>Indeed, our structural model implies that wages are increasing in firm-types (see Lemma 1), and simulations confirm that the model delivers wages that are also increasing in worker-types. Simulation studies in Bagger and Lentz (2015) indicate that the estimated “wage-types” from regressions like (2.37) captures “productivity-types” well provided wages are monotone in productivity-types.

in the polynomial specification for  $y(a, p)$ , see (2.30). In addition to wage data, we make (at the moment, limited) use of information on value added per worker for the firms in our data to help identify  $y(a, p)$ .

Specifically, we include the average cross sectional wage, the 20th and 80th percentiles in the distribution of estimated worker effects from (2.37), and the same percentiles in the distribution of estimated firm effects. The average log wage in our data is 5.2273, and the 20th and 80th percentiles from the distribution of worker and firm fixed effects are  $-0.2754$  and  $0.2418$  for the worker effects, and  $-0.0317$  and  $0.1631$  for the firm effects. Hence, the wage data exhibits economically meaningful heterogeneity in both the worker and the firm dimension. Recall that the grand mean of the firm effects are normalized to zero, and that the reported percentiles therefore suggests the distribution of firm fixed effects has a rather long right tail.

We also include the so-called co-worker correlation. This latter statistics is the correlation between a individual's worker fixed effect and the average worker fixed effects of all his co-workers, see Lopes de Melo (2018). In our data, the co-worker correlation turns out to be 0.4421, suggesting a relatively strong tendency for high (low) worker fixed effect individuals to work with other high (low) worker fixed effect individuals. In our model, such sorting patterns are primarily determined by differences in search behavior across worker-types, and these difference arise because the wage gains from reallocating from low- to high-type firms vary across the ability distribution, i.e. because of the modularity of the wage function  $w(a|p)$ . Since the wage offer distribution is endogenous, modularity of  $w(a|p)$  reflect, in part, the modularity of the match-level output function.<sup>34</sup>

Regarding value added data, we include the average firm-level labor-share, i.e. the ratio of a firm's annual wage bill to it's annual value added, as well as the employment-weighted correlation between value added per worker-hour and the average wage paid by the firm. As detailed in the description of our in section 2.3, we observe total annual value added from a firm's VAT account with the Danish tax authorities, and the number of hours per year per firm is then computed using the hours information in the spell data *before* the sample selection described in section 2.3 takes place. The average labor share is 0.8028 in our data, and the employment-weighted correlation between average value added per worker and the average wage paid by the firm is 0.1360. We note that the correlation between observed firm productivity, i.e. value added per hour, and the average wage paid is positive, but not particularly strong.

---

<sup>34</sup>Simulations confirm this to be the case, although we note that the Burdett-Mortensen wage setting game we apply here seems unable to deliver a submodular wage function, even when the output function is (strongly) submodular.

**Identification of the level of free search effort  $\underline{s}$ .** The free search parameter  $\underline{s}$  ensures all workers provide at least search effort  $\underline{s}$ , which first of all implies that the Diamond paradox is moot in the model. However, it also ensures that job search persists all the way to the top of the job ladder, which sustains an element of competition for workers at the higher rungs of the wage ladders. The competition manifest itself in the wage policies of firms, specifically in the variability of the firm fixed effects estimated from (2.37). The auxiliary statistics listed in the previous paragraphs relating to identification of  $y(a, p)$  are numerous enough to over-identify  $y(a, p)$ , and we use these degrees of freedom to argue that the difference between the 20th and 80th percentile in the firm fixed effect distribution identifies  $\underline{s}$ . Simulation studies and our estimation procedure confirms this logic.

**Identification of the disutility of job search effort function  $\zeta_0$  and  $\zeta_1$ .** The disutility from job search effort function  $\zeta(\cdot)$  is central to our analysis and has two parameters,  $\zeta_0$  and  $\zeta_1$ .  $\zeta_0$  determines the level of disutility, whereas  $\zeta_1$  governs the curvature, and therefore the substitution effect. If job search effort yields higher disutility at all effort-levels, job search effort will fall, both off- and on-the-job. Hence, higher disutility from job search, i.e. higher  $\zeta_0$ , reduces the outflow from nonemployment. For a given job destruction rate, this increases the steady state nonemployment rate, see (2.14). Hence, we take identification of  $\zeta_0$  from the inclusion of the average nonemployment rate in our five-year panel, which is 0.1180.<sup>35</sup>

Since  $\zeta_1$  governs the curvature of  $\zeta(\cdot)$ , it impacts differences in search effort between workers with different circumstances. Suppose the disutility function  $\zeta(\cdot)$  is extremely convex. In that case, workers with very different returns to search will choose, essentially, the same job search effort—because the the disutility function is effectively “kinked”. Hence, all else equal, an extremely convex disutility of search effort function, minimizes the variance of job durations, with all variation driven by variation in acceptance probabilities arising from wage dispersion. With less convex disutility, workers with different returns to search due to differences in ability and wages will make different choices regarding the search effort they provide. This increases the variance of job search effort, and therefore, all else equal, the variance of observed job durations. Based on that logic, we include the variance of the duration of the first job spell after non-employment in the vector of auxiliary statistics in order to identify  $\zeta_1$ . In our estimation panel, where job durations are necessarily censored after five years of duration at the latest, we measure this variance

---

<sup>35</sup>The nonemployment rate is computed among workers who, at some point during our estimation period, are observed working. Our data contains individuals who are never working, so the unconditional nonemployment rate exceeds 0.1180. Our model interprets workers who are never, or rarely, working as being of types with very little return to job search.

to be 1.5107, including both censored and non-censored jobs durations.<sup>36</sup>

**Identification of the hiring intensity cost function  $\chi_1$ .** Recall that the hiring cost function  $\chi(\cdot)$  is normalized such that  $\chi(1) = 1$ , which leaves one parameter to be identified, namely  $\chi_1$ , determining the curvature of the hiring cost function. This parameter is identified from the employment-weighted variance of firms' hiring rates. The logic is analogous to that applied above to argue identification of the curvature of the disutility of job search effort. We measure a firm's year  $t$  hiring rate as the ratio of new hires from November 28th, year  $t$  to November 27th, year  $t + 1$  to the number of employees in the firm on November 28th in year  $t$ , which allow us to compute firm-level hiring rates for 1999-2002.<sup>37</sup> The empirical employment-weighted variance of the hiring rate is 0.0827.

**Identification of the sunk entry cost  $C$ .** A high entry cost deters firm entry, which implies that the firms that do enter grow large. Hence, there is a positive relationship between the sunk entry cost parameter  $C$  and the average firm size in the economy. We consequently include the average firm size as an auxiliary statistic that we require our structural model to fit. A firm's size in year  $t$  is measured as the number of employees observed in the estimation data on November 28th in year  $t$ . The average firm size refers to the average taken across all firms and the five cross sections in our estimation panel ( $t = 1999, 2000, \dots, 2003$ ). We find an average firm size of 8.8414.

**Identification of the job destruction process  $\delta(a)$ .** Recall from (2.34) that we allow the job destruction process to be a function of worker ability. High-ability workers are over-represented high-wage workers, and we therefore use the empirical job destruction hazard conditional on workers' wages being below or above the median wage in the cross section. Specifically, consider all ongoing jobs at each annual cross section dates (November 28th), and compute the median wage in these jobs. Next, split the cross section sampled job spells according to whether the individual wage is above or below the median wage just computed. Finally, record the share of these jobs that end within a year because the worker transits to unemployment, for both below- and above median wage jobs. We find that the job destruction rate among jobs where workers are paid below the median wage is 0.2137 per year, whereas it drops to 0.0938 for jobs where workers are paid above the median wage. The model interprets the difference in job destruction rates as stemming from differences in the job destruction rate by worker ability.

<sup>36</sup>It is important to keep in mind that we replicate the data structure, and therefore this sampling of job durations, in the simulated data.

<sup>37</sup>We cannot compute the hiring rate for 2003, the final year in our estimation panel, because the 1999-2003 tax regime ends in December 31st, 2003.

**Identification of the job reallocation rate  $\mu$ .** It is straightforward to show that, in a “standard” model of on-the-job search where job offers arrive at exogenous rates to nonemployed and employed workers, such as e.g. Burdett and Mortensen (1998), the share of jobs in a cross section of workers that end with the worker making a transition to nonemployment, is bounded from below by  $1/2$ . With reallocation shocks this share may go below  $1/2$  because reallocations increases the prevalence of job-to-job transitions without increasing to the steady state mass of high-wage workers with high reservation wages. Hence, in this case, the reallocation rate is identified from the share of cross section jobs that end in a job-to-nonemployment transition—at least insofar as the empirical share falls below  $1/2$ .

We have not established a lower bound for the steady state share of jobs that end in a job-to-nonemployment transition in our more complicated model with worker and firm heterogeneity, and endogenous search effort choices on both sides of the market, but our simulation studies suggest that the model cannot easily deliver a share below  $1/2$  without the additional job-to-job transition from reallocation shocks. Empirically, the share is 0.4745, i.e. close to, but below  $1/2$ , thus providing identification of the reallocation rate  $\mu$ .

#### 2.5.4 Model fit

Having established identification of the model’s structural parameters, we turn to estimation. Before reporting the structural parameter values, we describe the estimated model’s fit to the auxiliary statistics just described. Table 2.6 reports the empirical and the simulated values for the auxiliary statistics used for estimation. In addition, Table 2.6 reports the diagonal elements in the Indirect Inference estimator’s weight-matrix  $\mathbf{\Omega}$ . All off-diagonal elements in  $\mathbf{\Omega}$  are set to zero. We put relatively high weight on average log wages and on the average firm size, and relatively little weight on the covariance between wages and hourly value added.

Overall, the fit reported in Table 2.6 is good. The over-weighting of average log wages ensures that the model accurately reproduces the level of wages. This is important as we applying the empirical tax function  $T(\cdot)$  to the simulated wages. The fit to the 20th and 80th percentile in the distribution of worker fixed effects is spot on as well. However, while the model accurately reproduces the 20th percentile in the distribution of firm fixed effects, it does not generate the long right tail observed in the distribution of firm fixed effects estimated using the real data. The estimated structural model therefore underestimates the variability in log wages that can be attributed to firm fixed effects.

Furthermore, the model simultaneously underestimates the correlation between an individual’s own worker fixed effect and her co-workers average worker fixed effect, and

overestimates the correlation between the average wage paid by firms, and their value added. Bagger and Lentz (2015), employing a random search model bearing similarities to the the present model, also underestimated the “worker, co-worker” fixed effect correlation to the same extent, and attributes this to workers sorting on other dimensions than wages (or productivity). We believe the overestimation of the correlation between wages and value added largely stem from the absence of measurement errors in the simulated wage and output measurements. For this reason, we put relatively little weight on the “firm wage, value added”-correlation in the estimation.

With respect to the fit of the auxiliary statistics that are particularly informative about the disutility of job search effort function  $\zeta(\cdot)$ , namely the nonemployment rate and the variance of durations of jobs initiated by a nonemployment-to-job transition, we see that the estimated model does an excellent job in reproducing the empirical magnitudes. This is reassuring given that the curvature of the disutility job search effort is the key determinant of the substitution effect in job search effort in response to income tax reforms. The fit to the employment-weighted variance of the hiring rate, on the other hand, leaves room for improvement. The estimated model under-predicts the variance of hiring rates by more than 40 percent.

The average firm size carries relatively high weight in the estimation, and as a consequence, the fit is excellent, as is the fit to the wage conditional job destruction hazards, even though these latter two moments carries normal weight in the estimation. Finally, the share of jobs ending in job-to-nonemployment transition, an auxiliary statistic included to identify the reallocation rate  $\mu$ , is fitted reasonably well, although the estimated model’s prediction exceeds 1/2, while the data, as noted above, has a share slightly below 1/2.

### 2.5.5 Structural parameter estimates

A subset of the structural parameter estimates are reported in Table 2.7, along with their asymptotic standard errors (see footnote 29). We do not report the parameter estimates of  $\delta_0$  and  $\delta_1$  that shape the job destruction process  $\delta(a)$ , nor of  $\boldsymbol{\varrho}$ , the parameters of the match output function  $y(a, p)$ , both which we instead render graphically in Figure 2.3.

Turning attention first on the estimates reported in Table 2.7, we note that, with an estimate of  $\zeta_1$  of 1.5257, the elasticity of job search effort disutility with respect to job search effort is 1.66, i.e. with slightly less curvature than a quadratic disutility function. This is consistent with the empirical results in Christensen et al. (2005), Kreiner et al. (2015), and Bagger and Lentz (2015), who all utilizes Danish data, and who all report approximately quadratic job search cost functions. We also note that  $\zeta_0$  is precisely estimated at 0.1011, that free search  $\underline{s}$  is estimated to be only 0.0212, and that the job reallocation rates  $\mu$  is estimated to be 0.1489. The latter three parameters, however, can

Table 2.6: Model fit

	Simulated	Data	Weight
Average log wage	5.2254	5.2273	100.00
Workers' share of value added	0.8407	0.8028	1.00
20th percentile, worker fixed effects	-0.2795	-0.2754	1.00
80th percentile, worker fixed effects	0.2405	0.2418	1.00
20th percentile, firm fixed effects	-0.0339	-0.0317	1.00
80th percentile, firm fixed effects	0.0606	0.1631	1.00
Corr(worker fixed effect, average co-worker fixed effects)	0.2681	0.4421	1.00
Corr(average firm wage, hourly value added)	0.2447	0.1360	0.10
Nonemployment rate	0.1289	0.1180	1.00
Variance, duration of the 1st job out of nonemployment	1.6097	1.5107	1.00
Variance, hiring rate (employment weighted)	0.0471	0.0827	1.00
Average firm size	8.9939	8.8414	10.00
Job destruction hazard if wage < median wage	0.1799	0.2137	1.00
Job destruction hazard if wage $\geq$ median wage	0.0918	0.0938	1.00
Share of jobs ending in job-to-nonemployment transition	0.5517	0.4745	1.00

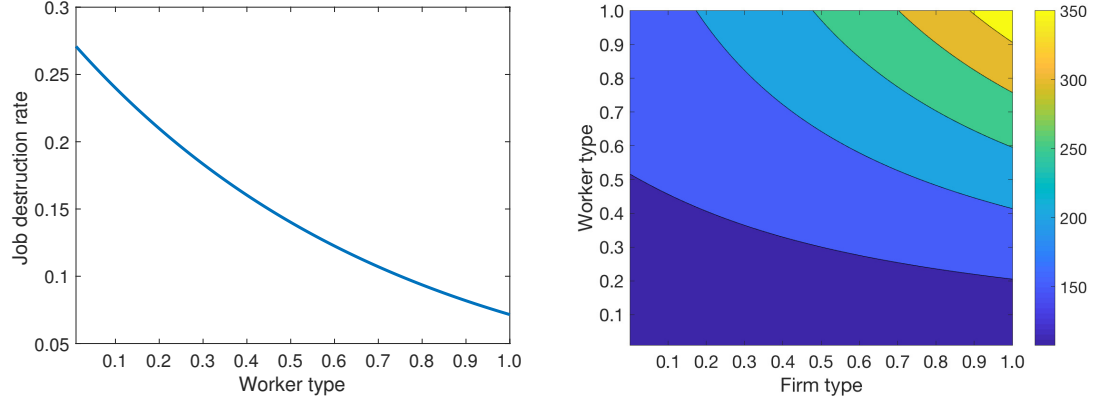
Note: The column labeled “weight” reports the diagonal elements in the weight-matrix  $\mathbf{\Omega}$ , see (2.35). Off-diagonal entries in  $\mathbf{\Omega}$  are set to zero.

Table 2.7: Structural parameter estimates

Parameter	Estimate
Disutility of job search effort $\zeta(s - \underline{s}) = \frac{\zeta_0}{1+1/\zeta_1} s^{1+1/\zeta_1}$	
Level, $\zeta_0$	0.1011 (0.0191)
Curvature, $\zeta_1$	1.5257 (0.2205)
Free search, $\underline{s}$	0.0212 (0.0010)
Reallocation rate (annual), $\mu$	0.1489 (0.0339)
Hiring intensity cost $\chi(v) = \frac{1}{1+1/\chi_1} s^{1+1/\chi_1}$	
Curvature, $\chi_1$	0.7244 (0.0390)
Sunk entry cost, $C$	5,082 (10.4026)

Note: Asymptotic standard errors are reported in parentheses.

Figure 2.3: Job destructions and match output



Panel A: Annual job destruction rate  $\delta(a)$

Panel B: Match output function  $y(a, p)$

Note: Match output is denominated in Danish Kroner per hour.

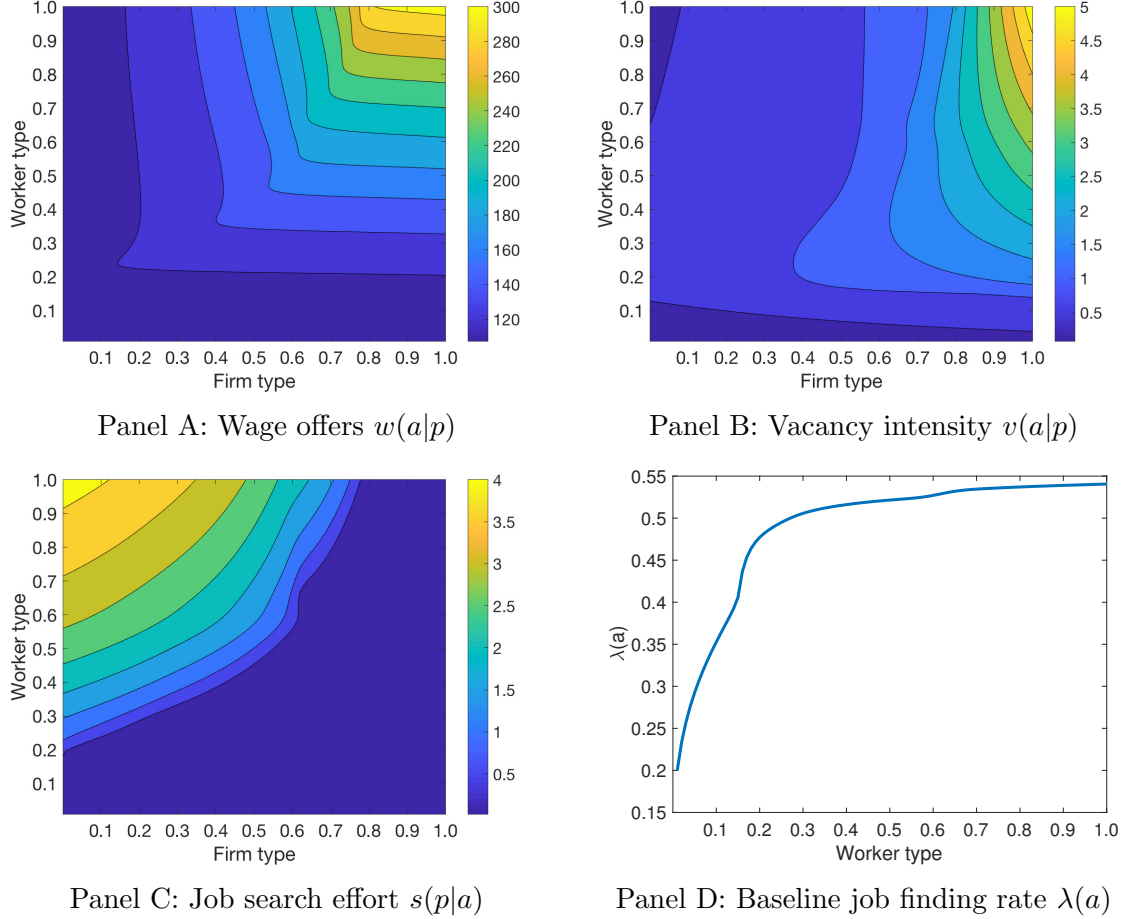
be interpreted only in proportion to the baseline job offer arrival rate  $\lambda(a)$ , an equilibrium object that we return to below. The final two parameter estimates reported in Table 2.7 refer to the firm-side of the model. Hiring cost is convex with the elasticity with respect to hiring effort being 2.38 and the sunk entry cost is estimated to be 5,082 Danish Kroner.

Panel A in Figure 2.3 plots the estimated annual job destruction process as a function of worker ability  $a$ . The job destruction rate exhibits a strong negative relationship with worker ability  $a$ . Indeed, a worker at the 25th percentile in the ability distribution has her job destroyed, on average, every fifth year, whereas a worker at the 75th percentile waits, on average, 10 years between job destruction events. Panel B in Figure 2.3 is a contour plot of the estimated match output function  $y(a, p)$ . By assumption,  $y(a, p)$  is increasing in both arguments, but it is evident that the estimated match output function is supermodular: The increase in output from reallocating a high-ability worker from a low- to a high-productive employer is greater than the increase realized by the same reallocation of a low-ability worker. Likewise, the increase in output from replacing a low-ability worker with a high-ability worker is greater for firms with higher productivity. This completes the description of the model's exogenous components.

The estimated model favors high-type workers in two important ways. First, high-type workers face lower job destruction rates than low-type workers. All else equal, this ensures that, on average, high-type workers are matched with better firm-types than low-type workers. It also increases the return to job search for high-type workers vis-a-vis low-type workers. Second, the match output function is increasing in  $a$  by construction, and in addition, is estimated to be supermodular. Insofar as the supermodularity is carried



Figure 2.4: Key equilibrium objects



Note: Wages are denominated in Danish Kroner per hour.

over to the equilibrium wage function, and provided income effects, and any countervailing equilibrium effects running through firms hiring intensity choices  $v(a|p)$  or the equilibrium baseline job offer arrival rate  $\lambda(a)$  are relatively small, this will tend to further widen the gap in returns to job search effort between high- and low-type workers.

Figure 2.4 depicts a set of key equilibrium objects relevant for understanding the implications of the estimated model for the equilibrium allocation of labor. Panel A plot the contours of the estimated equilibrium wage function  $w(a|p)$ . Of course, as established in Lemma 1, wages are strictly increasing in  $p$ , but, in addition, higher-type workers almost always reap greater wage gains by moving from a low- to a high-productive firm than do workers of lower ability. Unlike the estimated match output function, the wage function is,

however, not globally supermodular.<sup>38</sup> Panel B plots a contour plot of firms' equilibrium hiring intensity as a function of worker- and firm-types. Consistent with Lemma 2, hiring intensity  $v(a|p)$  is always increasing in  $p$  for a given  $a$ , but we see that hiring intensity  $v(a|p)$  need not be increasing in  $a$  for a given  $p$ , although there is a very strong tendency for this to be the case.

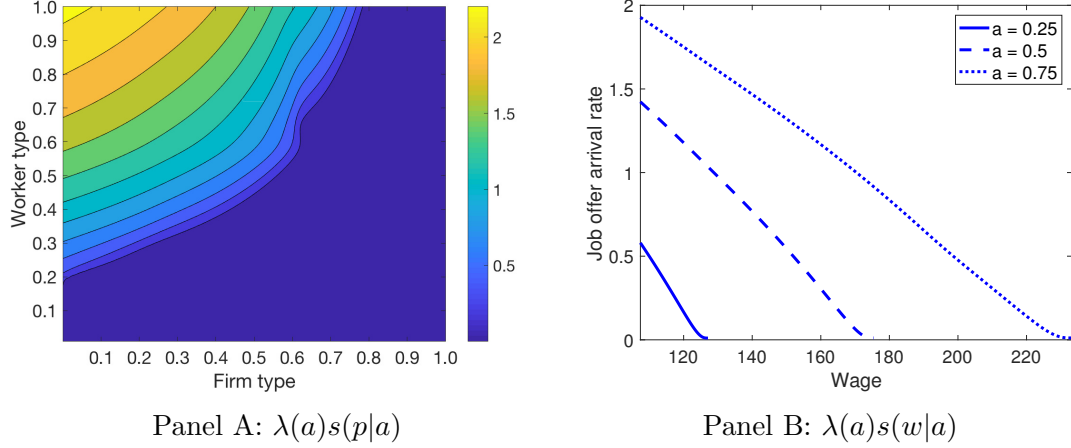
Turning to the key decision margin of workers—job search effort—Panel C in Figure 2.4 contains a contour plot of workers' optimal job search effort as a function of worker- and firm-types. As stipulated by Lemma 1, job search effort is strictly decreasing in  $p$  for each  $a$  with optimal search effort in the most productive firm being  $\underline{s}$  for all  $a$ . Moreover, we see that, at any firm-type  $p$ , higher-type workers exert more job search effort than lower-type workers. Hence, the equilibrium implies higher return to search for high-type workers.

As noted further above, the magnitude of job search effort, free search, reallocation shocks must be interpreted in proportion to the relevant baseline job offer arrival rate  $\lambda(a)$ . Panel D in Figure 2.4 plots  $\lambda(a)$  as a function of  $a$ . We see that  $\lambda(a)$  is strictly increasing and concave function in  $a$ . The labor market for the least able workers receive job offers at an annual rate of 0.2 per unit of search, whereas, at the other end of the ability spectrum, high-ability workers receive offers at annual rate 0.55 per unit of search, emphasizing again the estimated equilibrium is such that return to job search effort is higher for high ability workers.

Finally, Figure 2.5 puts Panels C and D from Figure 2.4 together and plots the annual job offer arrival rates resulting from on-the-job search effort implied by the estimated model. Panel A plots the job offer arrival rate as a function of firm-type,  $\lambda(a)s(p|a)$  by way of a contour plot. Because both  $s(p|a)$  and  $\lambda(a)$  are increasing functions of  $a$ , conditional on the type of the employing firm, workers of higher ability will, in equilibrium, receive more job offers than workers of lower ability. Panel B in Figure 2.5 plot the job offer arrival rate for workers of type  $a = 0.25$ ,  $a = 0.50$  and  $a = 0.75$ , respectively, as a function of the wage, i.e.  $\lambda(a)s(w|a)$  where  $w = w(a|p)$ . We see that, at any wage rate, more able workers receive more job offers. Indeed, at the lowest wage, equilibrium behavior implies that a type  $a = 0.75$  worker is four times more likely to receive a job offer than a type  $a = 0.25$  worker. We stress that these differences arise endogenously, and that they have implications for the equilibrium allocation of labor, a subject we turn to next.

<sup>38</sup>We are not entirely sure what causes this feature of the equilibrium wage function, and cannot at the time of writing rule out that it is caused by numerical problem in computing the equilibrium wage function.

Figure 2.5: Estimated annual job offer arrival rates by worker- and firm-type



Note: Panel A plots the estimated job offer arrival rates as a function of firm-type  $p \in [p_0, 1]$  and the worker-type  $a \in [a_0, 1]$ . Panel B plots the estimated job offer arrival rates as a function of wage  $w = w(a|p) \in [b, \bar{w}(a)]$ .

### 2.5.6 The equilibrium allocation of labor

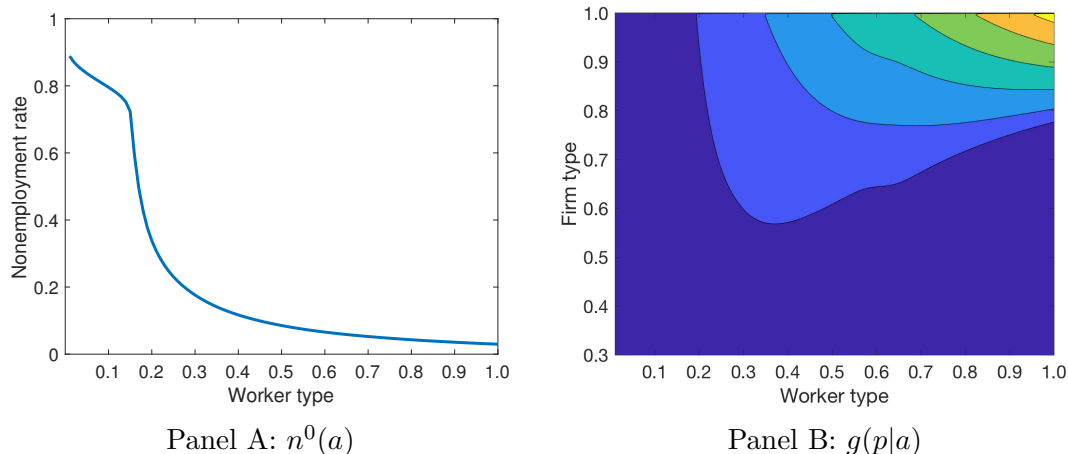
The allocation of labor is represented by the worker-ability specific nonemployment rates,  $n^0(a)$  given by (2.14), and the worker-ability specific distributions of workers across wage rates  $w$ ,  $G(w|p)$  given by (2.16). Note that, since  $w = w(a|p)$  is a one-to-one function of  $p$ , the distribution of worker-types across firm types is simply given as  $G(p|a) = G(w(a|p)|a)$ .

We have already seen from Table 2.6 that the estimated model implies an empirically reasonable aggregate nonemployment rate  $n^0 = \int_0^1 n^0(a) dH(a)$  around 0.13 among those workers ever observed in employment over the 5 years of simulated data. However, the aggregate nonemployment rate masks heterogeneity across worker-types. Panel A in Figure 2.6 plots the estimated steady state nonemployment rates as a function of worker ability  $a$ . We note that the estimated model implies that the workers in the lowest quintile of the ability distribution face very high nonemployment rates, in excess of 20 percent, while workers in the top quintile of the ability distribution has nonemployment rates around 2 percent.<sup>39</sup>

Panel B in Figure 2.6 illustrates the conditional densities  $g(p|a)$  as a function of  $a$ . We see a clear pattern whereby high-type workers tend to cluster in high-type firms to a much higher extent than low-type workers. In that sense, the equilibrium allocation exhibits traits of positive assortative matching, a finding that is consistent with Bagger and Lentz

<sup>39</sup>Workers at the very bottom of the ability distribution are effectively estimated to be nonparticipants. Many of these workers will not be observed in employment during a 5-years simulation window, and will therefore not contribute to the nonemployment rate of 0.13 reported in Table 2.6. Our data, of course, also contains nonparticipants.

Figure 2.6: The equilibrium allocation of labor



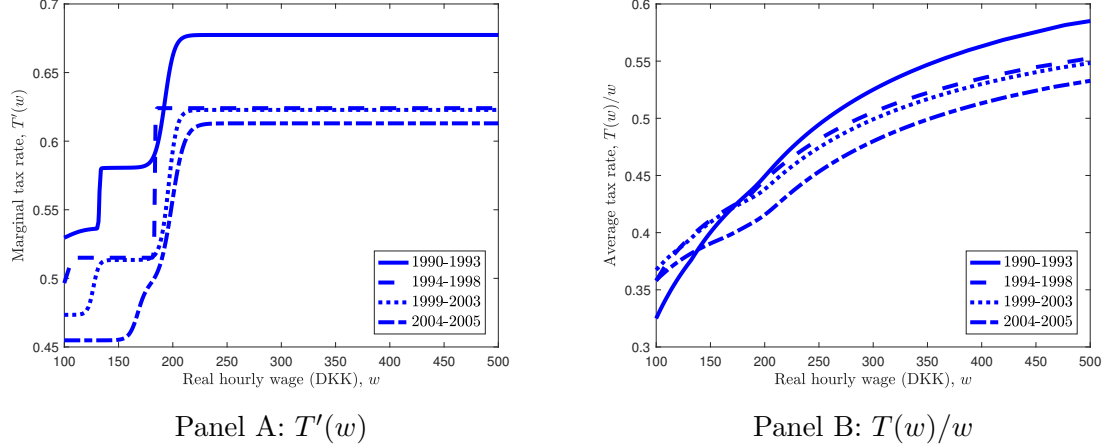
(2015). This equilibrium sorting results come about because high ability workers have higher returns to job search effort, due to their lower unemployment risk, see Panel A in Figure 2.3, and the super-modularity of the match output function, see Panel B in Figure 2.3. The endogenous response to the higher returns to job search effort result in higher job offer arrival rates, see Panel C in Figure 2.5, which implies that high-ability workers climb the wage (or firm productivity) distribution faster than low-ability workers. The remainder of the paper analyzes and quantifies the impact of the income tax schedule on the equilibrium allocation of labor.

## 2.6 Equilibrium tax reform evaluations

We initiate our analysis of the impact of labor income taxation on the allocation of labor by employing the estimated structural model to conduct equilibrium evaluations of a series of Danish income tax reforms during 1990-2005. The reforms under consideration took place in 1990, 1994, and 2004 and the context in which they were devised and implemented is described in section 2.4.1. Figure 2.7 shows the marginal tax functions (Panel A) and the average tax functions (Panel B) for the four tax regimes we consider, including the 1999-2003 regime used for estimation. Inspection of Figure 2.7 reveals that each of the implemented changes in the income tax system during 1990-2005 (almost) uniformly flattened the tax function, i.e. lowered the marginal rate at all wage rates. At the same time, the average tax rate was increased at the bottom of the wage distribution, and lowered everywhere else.

The equilibrium evaluations are conducted as comparative statics between the steady state equilibria associated with each of the income tax regimes. The structural param-

Figure 2.7: Danish tax regimes 1990-2005: Marginal and average tax functions



Note: Panel A plots the marginal tax function  $T'(w)$  and Panel B plots the average tax function  $T(w)$ .

ters are held fixed throughout. Our interest centers on aggregate and worker-type specific steady state, i.e. long run, nonemployment rates and expected labor income. The worker-type specific nonemployment rate  $n^0(a)$  is given by (2.14), with the aggregate nonemployment rate given by  $\int_0^1 n^0(a) dH(a)$ . The worker-type  $a$  specific expected labor income is given by

$$LI(a) = [1 - n^0(a)] \int_0^1 w(a|p) dG(w(a|p)|a), \quad (2.38)$$

where  $G(w|a)$  is given by (2.16), and aggregate labor income is  $LI = \int_0^1 LI(a) dH(a)$ .

The tax reforms also impact the government budget. As mentioned above, our current equilibrium computation algorithm does not feature lump-sum transfers to balance the government budget. Instead, to at least provide context for the reported effects on the equilibrium labor allocation, we also report the change in government budget associated with each of the tax regime comparisons.

Table 2.8 reports the percentage change in nonemployment  $n^0$ , labor income  $LI$ , and the government budget  $B$  between each of the 1990, 1994, and 2004 Danish income tax regimes, and the baseline 1999 regime. Hence, in Table 2.8, the 1999 tax regime is always the baseline. This means that the numbers reported in the 1990-1993 panel evaluates the effect of changing the 1999-2003 regime (back) to the 1990-1993 regime, while the numbers in the 2004-2005 panel evaluates the effect of the actual reform 2004 reform, i.e. changing the 1999-2003 regime to the 2004-2005 regime. Perhaps confusingly, a negative number for, say, labor income in the 1990-1993 panel therefore means that the allocation *improved* between the 1990-1993 regime and the 1999-2003 regime, while a negative number for labor

income in the 2004-2005 panel means that the allocation of labor *deteriorated* between the 1999-2003 regime and the 2004-2005 regime.

We consider the aggregate response (rows labeled “All” in Table 2.8), as well as the responses at the 25th, 50th, and 75th percentile of the ability distribution (rows labeled 0.25, 0.50, and 0.75, respectively). For each variable, we consider the total effect, comprising both the substitution effect and the income effect, as well as the substitution effect in isolation. Hence, the difference between pairs of columns labeled “Total” and “Subst” in Table 2.8 is the income effect.<sup>40</sup> Finally, we consider both the partial equilibrium effect (labeled “PE”) and full equilibrium effect (labeled “FE”). The partial equilibrium effect holds firm behavior and equilibrium variables constant, allowing only workers to respond to changes in the income tax code.

**Long run nonemployment.** Relative to the 1999-2003 tax regime, the 1990-1993 tax regime involves a 4.9 percent higher steady state nonemployment rate when accounting for equilibrium effects. The corresponding partial equilibrium effect is much larger at 7.4 percent. Income effects are modest at the aggregate level. In partial equilibrium, nonemployment increases 7.9 percent due to the substitution effect. This causes an increase in the marginal value of consumption, and the resulting income effect in workers’ job search effort reduces the total effect slightly to 7.4 percent. The full equilibrium adjustment is such that the income effect in the aggregate, albeit small, works in the same direction as the substitution effect. If we look at the effect within the ability distribution, we see that workers at the bottom of the ability distribution experience the largest increase in nonemployment. The differences across ability levels is especially pronounced in full equilibrium. We also note that the positive aggregate income effect in full equilibrium is driven by equilibrium adjustments at the bottom of the ability distribution—towards the top of the ability distribution, the income effect is negative, dampening the substitution effect.

Similar results, although with smaller magnitudes are found when comparing the 1994-1998 regime to the baseline 1999-2003 regime. Although full equilibrium income effects are still positive towards the bottom of the ability distribution, the aggregate income effect is negative, but still small. In full equilibrium, aggregate nonemployment is 1.1 percent higher in the 1994-1998 regime than in the 1999-2003 regime. The distinction between partial and full equilibrium adjustments continues to be quantitatively important. In comparison to the 1999-2003, the 2004-2005 regime involves 1.4 percent lower nonemployment in full equilibrium, with the full equilibrium substitution effect being a 2.6 percent lower nonemployment rate. In the case of the 2004-2005 regime, income effects become economically significant, often reducing the substitution effect by up to 50 percent.

---

<sup>40</sup>The substitution effect holds the marginal utility of consumption constant, see section 2.2.2.2.

Table 2.8: Evaluating the 1990, 1994, and 2004 Danish income tax reforms for nonemployment, labor income and the government budget

		$\frac{\Delta n^0}{n^0} \times 100\%$				$\frac{\Delta LI}{LI} \times 100\%$				$\frac{\Delta B}{B} \times 100\%$				
Comparison	tax regime	Ability	Total		Subst		Total		Subst		Total		Subst	
			PE	FE	PE	FE	PE	FE	PE	FE	PE	FE	PE	FE
1990-1993		All	7.4	4.9	7.9	4.8	-1.0	-1.8	-1.0	-1.9	-1.6	-2.7	-1.5	-2.5
		0.25	10.3	9.9	7.8	5.7	-3.1	-3.9	-2.4	-2.2	-22.9	-24.1	-20.2	-18.9
		0.50	9.3	5.3	9.3	5.4	-1.0	-1.7	-1.0	-1.8	-2.4	-2.8	-2.4	-2.8
		0.75	8.1	3.5	10.4	5.6	-0.6	-1.3	-0.8	-1.8	2.3	1.6	2.0	0.9
1994-1998		All	1.7	1.1	1.9	1.2	-0.3	-0.5	-0.3	0.5	-0.3	-0.6	-0.3	-0.6
		0.25	3.3	3.2	2.9	2.6	-1.0	-1.2	-0.9	1.0	-5.7	-6.1	-5.2	-5.3
		0.50	1.8	0.8	1.7	0.7	-0.2	-0.3	0.9	0.8	-0.6	-0.6	-0.6	-0.5
		0.75	1.8	0.6	2.3	1.1	-0.1	-0.4	-0.2	-0.5	0.7	0.5	0.6	0.3
2004-2005		All	-1.7	-1.4	-3.6	-2.6	0.2	0.3	0.4	0.7	-5.1	-5.0	-4.6	-4.4
		0.25	-1.4	-1.7	-2.6	-2.4	0.4	0.8	0.8	1.1	-4.0	-3.2	-2.7	-2.4
		0.50	-2.9	-2.3	-5.4	-4.0	0.3	0.5	0.6	1.0	-4.8	-4.7	-4.3	-4.1
		0.75	-1.9	-1.3	-4.5	-2.8	0.1	0.0	0.3	0.5	-4.4	-4.5	-4.1	-4.0

Note: The percentage changes are relative to the 1999-2003 income tax regime. Hence, positive (negative) numbers in the 2004-2005 panel indicates that the variable increased (decreased) going from the 1999-2003 regime to the 2004-2005 tax regime. Positive (negative) numbers in the 1994-1998 and 1990-1993 panels, however, indicate that the the variable decreased (increased) as a result of going from the 1994-1998 (or 1990-1993) tax regime to the 1999-2003 regime.

Overall, each of the 1990, 1994, and 2004 Danish income tax reforms lead to a reduction in nonemployment. The effect reported in Table 2.8 implies that the tax reforms implemented during 1990-2005 in Denmark reduced steady state nonemployment by approximately 6 percent. However, towards the bottom of the ability distribution, the reduction was larger, around 13 percent at the 25th percentile in the ability distribution, while the effect is smaller for workers with higher ability, around 3% among workers at the 75th percentile in the ability distribution. The income effect reduces the effect for high ability workers in particular.

**Long run labor income.** Consider first the evaluation of the 1990-1993 tax regime relative to the baseline 1999-2003 regime in the first panel of Table 2.8. Table 2.8 shows that the overall effect amounts to a 1.8 percent drop in labor income, accounting for equilibrium adjustments. The partial equilibrium response is a 1 percent drop in labor income, again highlighting the importance of modeling the equilibrium response to tax reforms. On the aggregate, the income effects are small. However, we again find heterogenous effects across the ability distribution with low ability workers experiencing a much larger reduction in their labor income than workers of higher ability.

Turning now to the comparison between the 1994-1998 regime relative to the baseline 1999-2003 regime, we obtain a smaller increase in aggregate labor income of 0.5 percent accounting for equilibrium adjustment. The partial equilibrium response indicates a 0.3 percent decrease in labor income. Accounting for the income effects in the comparison between the 1994-1998 regime and the 1999-2003 regime is quantitatively important. The substitution effect indicate a 0.5 increase in labor income, implying that the income effect reverses the sign of the predicted labor income response. The labor income response again exhibits heterogeneity across the ability distribution.

Finally, consider the comparison of labor incomes between the 2004-2005 tax regime and the baseline 1999-2003 regime, i.e. the third panel in Table 2.8. Going from the 1999-2003 tax regime to the 2004-2005 regime involves a 0.3 percent long run gain in aggregate labor income, when equilibrium adjustments on the labor market are taken into account. These equilibrium adjustments are, however, small, with the partial equilibrium effect being a 0.2 percent increase. The relatively small total effect masks a large substitution effect increasing steady state labor income by 0.7 percent. Again, we see evidence of heterogenous responses throughout the ability distribution.

The overall long run effect on labor income of the tax reforms implemented in Denmark between 1990 and 2005 is an approximate 2 percent increase, with larger effects (an approximate 4.5 percent long run increase) at the 25th percentile of the ability distribution, whereas for the more able workers at the top of the ability distribution, smaller effects



materialize (an approximate 1 percent increase).

**Long run government budget.** The third set of columns in Table 2.8 present the percentage change in the government budget, comparing the 1990-1993, the 1994-1998, and the 2004-2005 tax regimes to the baseline 1999-2003 regime. The reported figures imply that, in full equilibrium and accounting for income effects, the government budget increased 2.7 percent when going from the 1990-1993 regime to the 1999-2003 regime. If we consider the effect of going from the 1990-1993 to the 1999-2003 regime on the ability-specific contributions to the government budget, we see a large increase in tax revenue at the bottom of the ability distribution, with the effect tapering off as we move up through the distribution.

The budgetary implications of going from the 1994-1998 regime to the 1999-2003 regime are quite small, and we shall not comment further on them here, except to note that, accounting for equilibrium effects, the government budget increased 0.6 percent as a result of going from the 1994-1998 regime to the 1999-2003 regime. Keeping in mind that the average tax rate was reduced everywhere, except at quite low wage levels, it is interesting to note that government budget nonetheless increased as a result, perhaps suggesting the Danish labor market operated to the right of the peak of the Laffer-curve prior to 1999.<sup>41</sup>

Finally, consider the budgetary implications of going from the 1999-2003 regime to the 2004-2005 regime. From Table 2.8, we see that this reform leads to a rather large declines in government revenue, of 5 percent in full equilibrium, with the partial equilibrium effect, and the pure substitution effects being of the same magnitude. Considering Panel B in Figure 2.7, we note that the going from the 1999-2003 regime to the 2004-2005 regime involves uniformly reducing the average tax rates at all wage levels. The improved allocation is simply not enough to compensate for the associated mechanical loss. Given the sizeable budgetary implications involved in the 2004 income tax reform, one might speculate that allowing for lump sum transfers to balance the budget may impact the conclusions we reached for this reform regarding nonemployment and labor income.

**Summing up.** A series of income tax reforms in Denmark implemented in 1994, 1999, and 2004 improved the equilibrium allocation of labor, reducing unemployment by about 6 percent, and increasing labor income by about 2 percent, with workers of low ability gaining the most. Our analysis does not balance the government budget, and overall, the tax reforms during 1990-2005, lead to a 2 percent drop in government revenue. Our analysis

---

<sup>41</sup>In our model, the “Laffer effect” is operates through labor allocation adjustments. By lowering taxes, the government sustains a mechanical loss in revenue, but the behavioral effects on workers’ job search behavior improves the allocation of labor, leading to increases in government revenue. The behavioral gain may dominate the mechanical loss.

has shown that the key ingredients in our model—two-sided heterogeneity, taking account of possible income effects, and equilibrium adjustments—are quantitatively important for understanding the long run labor market adjustments to the implemented income tax reforms. Appendix A.5 provides a complementing analysis of counterfactual linear income tax functions to assess the impact of progressivity on the equilibrium allocation of labor.

## 2.7 The long run elasticity of taxable labor income

The evaluation exercises in the previous section demonstrated that, when analyzed through the lens of our structural model, the long run allocation of labor respond in economically meaningful ways to the incentives provided by the income tax system. The measured response operates through the job search effort choices of workers, and the hiring intensity choices and wage policies of firms, or put succinctly, through labor mobility.

Following Feldstein (1995, 1999), the broader public economics literature typically summarizes economic responses to taxation in the elasticity of taxable income, or more recently the elasticity of taxable labor income. The shift in focus towards taxable labor income reflects that labor income is less prone to evasion and avoidance, which means that behavioral responses can be understood within standard models of the labor market, see e.g. Chetty, Friedman, Olsen, and Pistaferri (2011a). From a theoretical point of view, the elasticity of taxable labor income comprises all behavioral responses to changes in the tax system of relevance for labor income, including hours responses, and effort provision in various dimensions. Empirically, Kleven and Schultz (2014b) report estimates of the taxable labor income elasticity in Denmark ranging between 0.05 and 0.12.<sup>42</sup>

Estimates of the taxable labor income elasticity, such as the ones reported in Kleven and Schultz (2014b), are often obtained using variation in marginal tax rates induced by tax reforms as a sources of exogenous variation in incentives across individuals. As discussed in Kreiner et al. (2015), such an approach is, however, unlikely to capture the search induced labor allocation effects. Indeed, Kreiner et al. (2015) uses a partial equilibrium job search model to estimate the long run elasticity of taxable labor income for Denmark, accounting for steady state labor allocation responses, at 0.30. The model used for estimation in Kreiner et al. (2015) does not feature equilibrium responses, does not explicitly account for worker heterogeneity, and does not allow for income effect in the job search effort responses of workers. Our model incorporates all these features. It is therefore of interest to revisit the elasticity of taxable labor income within the context of our richer model

---

<sup>42</sup>For the US, the elasticity of taxable income is typically estimated to be around 0.25, see Saez, Slemrod, and Giertz (2012a).

Table 2.9: Long run taxable labor income elasticities

Ability	Total		Subst	
	PE	FE	PE	FE
All	0.069	0.123	0.091	0.169
0.25	0.217	0.289	0.241	0.297
0.50	0.068	0.110	0.093	0.156
0.75	0.031	0.081	0.060	0.142

To proceed, recall that, within our model, taxable labor income  $LI = \int_0^1 L(a)dH(a)$ , where  $LI(a)$  is given by (2.38).  $LI(a)$  depends on the equilibrium allocation of labor across wage levels  $G(w|a)$ , see (2.16). The labor allocation, in turn, depends on workers' search effort which may be distorted by income taxation. Following Kreiner et al. (2015), we compute the elasticity of taxable labor income for type- $a$  workers, by recording the percentage change in their labor income  $LI(a)$  following a uniform 1 percent increase in the marginal net-of-tax rate  $1 - T'(w)$ , keeping  $b - T(b)$  constant.<sup>43</sup> Let  $\epsilon(a)$  denote the computed elasticity of taxable labor income for type- $a$  workers. The aggregate elasticity of taxable labor income is then  $\epsilon = \int_0^1 \epsilon(a)dH(a)$ .

Table 2.9 reports the partial and full equilibrium elasticities implied by the estimated structural model, both aggregate, and for workers at the 25th, 50th and 75th percentile in the ability distribution. Table 2.9 reports both the total effect, comprising both a substitution and an income effect, as well as elasticity estimates based purely on the substitution effect.

The results in Table 2.9 echoes a common theme running through this paper, namely that the distinctions between partial and full equilibrium, income and substitution effects, as well as worker heterogeneity are quantitatively important. We estimate the full equilibrium mobility-based elasticity of taxable labor income, comprising both substitution and income effects, to be 0.123, a substantially lower elasticity estimate than that reported by Kreiner et al. (2015). When we shut down equilibrium adjustments and consider only the substitution effect to facilitate closer correspondence to the modeling framework employed in Kreiner et al. (2015), we obtain an elasticity of 0.091.<sup>44</sup> In full equilibrium, the uncompensated elasticity of taxable labor income is larger, at 0.169. Finally, from Table 2.9, we note a clear pattern, whereby workers of lower ability have higher taxable labor income

<sup>43</sup>That is, we change the net-of-tax rate faced by workers from  $1 - T'(\cdot)$  to  $1.01 \times [1 - T'(\cdot)]$ .

<sup>44</sup>It is at present not clear to us why the difference between the two estimates arises. We have estimated the Kreiner et al. (2015) model on our data and obtained results similar to those they report, ruling out differences in the data we use. Kreiner et al. (2015) do, however, use a different tax function than we do, but we have yet to ascertain that differences in the tax function are responsible for the different estimates.

elasticities. Indeed, workers at the 25th percentile in the ability distribution have taxable labor elasticities close to the 0.30 reported in Kreiner et al. (2015), whereas workers at the 75th percentile have three times lower elasticities, at 0.081.

### 2.7.1 Decomposing the long run elasticity of taxable labor income

Our rich structural model allow us to further decompose the elasticity of taxable income to gain further insights into the behavioral (mobility) responses to income taxation. For this purpose, let  $LI(a)$  and  $\hat{LI}(a)$  be expected taxable incomes of a type- $a$  worker under two different income tax regimes,  $T(\cdot)$  and  $\hat{T}(\cdot)$ . Also, let  $n^0(a)$  and  $\hat{n}^0(a)$ ,  $G(p|a)$  and  $\hat{G}(p|a)$ , and  $w(a|p)$  and  $\hat{w}(a|p)$  be the nonemployment rates, wage distribution, and wage policies under the  $T(\cdot)$  and  $\hat{T}(\cdot)$  regimes, respectively. Given the expression for the expected taxable income of a type- $a$  worker, (2.38), the following decomposition of the change in taxable labor income between the two tax regimes  $T(\cdot)$  and  $\hat{T}(\cdot)$  is of interest:

$$\begin{aligned}
LI(a) - \hat{LI}(a) &= \underbrace{[n^0(a) - \hat{n}^0(a)] \int_0^1 \hat{w}(a|p) d\hat{G}(p|a)}_{\text{Extensive margin allocation}} \\
&+ \underbrace{[1 - n^0(a)] \int_0^1 \hat{w}(a|p) [\hat{g}(p|a) - g(p|a)] dp}_{\text{Intensive margin allocation}} \\
&+ \underbrace{[1 - n^0(a)] \int_0^1 [\hat{w}(a|p) - w(a|p)] dG(p|a)}_{\text{Wage policy margin}}. \tag{2.39}
\end{aligned}$$

Equation (2.39) states that the change in taxable labor income between the two tax regimes  $T(\cdot)$  and  $\hat{T}(\cdot)$  can be decomposed into responses along three different margins. First, a change in income taxation brings about a change in the equilibrium nonemployment rate, an extensive margin allocation effect. Second, a change in income taxation also shifts the equilibrium distribution of workers across firms, denoted the intensive margin allocation effect. Third, a change in income taxation induces changes in the equilibrium wages offered by firms, an effect we denote the wage policy margin.<sup>45</sup>

With (2.39) in hand, we can decompose the elasticity of taxable labor income estimates provided in Table 2.9. We focus here exclusively on the decomposition of the elasticity accounting for equilibrium adjustments, which was found to be 0.123 at the aggregate in

<sup>45</sup>The labor income decomposition (2.39) is carried out such that the extensive margin allocation is evaluated at the intensive allocation and the wage policy under  $\hat{T}(\cdot)$ , and the intensive allocation margin is evaluated at the extensive margin allocation under  $T(\cdot)$ , and the wage policy under  $\hat{T}(\cdot)$ . The wage policy margin is evaluated at the allocation under  $T(\cdot)$ . The decomposition is robust to the order in which we evaluate the three terms.

Table 2.10: Decomposing the full equilibrium elasticity of taxable labor income

	All	$a = 0.25$	$a = 0.50$	$a = 0.75$
Elasticity of taxable labor income	0.123	0.289	0.110	0.081
Extensive margin allocation	46%	73%	32%	15%
Intensive margin allocation	12%	2%	15%	21%
Wage policy margin	42%	24%	53%	64%

Table 2.9.<sup>46</sup> The decomposition is reported in Table 2.10, where we see that 46 percent of the aggregate taxable labor income elasticity can be ascribed to reallocations on the extensive margin, whereas 12 percent can be ascribed to reallocations on the intensive margin. Adjustments on the wage policy margin account for the remaining 42 percent of the estimated elasticity of taxable labor income.

When we apply the decomposition (2.39) to taxable labor income elasticities across the ability distribution an interesting pattern emerge. Adjustments along the extensive margin are responsible for the bulk of the measured taxable labor income elasticity at the bottom of the ability distribution. In Table 2.10, extensive margin adjustments account for 73 percent of measured elasticity among workers at the 25th percentile of the ability distribution. This share declines monotonically to 15 percent once we consider workers at the 75th percentile of the ability distribution. By implication, the opposite pattern emerges when we consider adjustments along the intensive margin, i.e. changes in  $G(p|a)$  and wage policies, i.e. changes in  $w(a|p)$ . The relative importance of both of these two margins of adjustment increases monotonically in the ability of workers.

## 2.8 A Pareto optimal tax reform

We have throughout the paper refrained from discussing normative issues, i.e. optimal income taxation. Our model has an exclusively positive focus and is designed to capture the enormous amount heterogeneity present in labor market data, and to deliver empirically realistic behavior, including wage distributions. As result, the model describes an economy that is likely rife with inefficiencies arising from congestion and thick market externalities. Such a model does not lend itself well to the analysis of optimal, i.e. welfare maximizing, policies for the simple reason that it is not possible to disentangle the extent to which a particular tax schedule is “optimal” because it happens to correct inefficiencies particular to the Danish 1999-2003 labor market under consideration here, or because it

<sup>46</sup>Decompositions of the partial equilibrium elasticities yield qualitatively similar results and are available upon request.

optimally trades efficiency off with equity.<sup>47</sup>

In this last section of the paper, we do, however, take a step towards analyzing a notion of optimal taxation. Specifically, inspired by Werning (2007), Blundell and Shephard (2012) and Hosseini and Shourideh (2017), we use the estimated structural model to construct a Pareto optimal income tax reform. That is, we consider whether—taking the status quo in terms of the actual tax system and any inefficiencies that may be present in the baseline economy as given—it is possible to reform the income tax system in such a way that the government tax revenue increases while all workers are made (weakly) better off in terms of steady state expected utility.

To state the problem we solve formally, let  $\bar{W}(a)$  be the expected steady state utility of a type- $a$  worker under the current tax system. We then solve

$$\max_{T(\cdot)} \int_0^1 \left\langle n^0(a)[T(b) - b] + [1 - n^0(a)] \int_b^{\bar{w}(a)} T(w) dG(w|a) \right\rangle dH(a) \quad (2.40)$$

subject to

$$n^0(a)\psi(c(a), s(b|a)) + [1 - n^0(a)] \int_b^{\bar{w}(a)} \psi(c(a), s(w|a)) dG(w|a) \geq \bar{W}(a) \quad (2.41)$$

for all  $a \in [0, 1]$ . At present, when searching for a Pareto optimal tax regime  $T(\cdot)$  that solves (2.40) subject to (2.41), we retain the functional form of the actual tax system, but allow for different parameter values.<sup>48</sup>

The resulting Pareto improving tax system is superimposed on the actual 1999-2003 Danish income tax system in Figure 2.8. The reform increases government revenue by 0.68 percent while leaving no workers worse off. In fact, workers' steady state utility, on average, increases slightly by 0.07 percent. From Figure 2.8 we see that the Pareto optimal income tax regime reduces the marginal tax rate at low wage levels, and increases it slightly at all higher wage levels. At the same time, however, the average tax rate is increased at low wages, and reduced at higher wage levels.

It is interesting to observe that the Pareto optimal tax schedule captures some qualitative features of the actual tax reforms implemented in Denmark during 1990-2005, and analyzed in section 2.6. In particular, tax rates at the very bottom of the wage distribution are increased, while they are lowered elsewhere. The Pareto optimal tax reform, however, combines these shifts in the tax function with a steeper marginal tax function, reducing marginal taxes for low wages and increasing it slightly elsewhere except at the very top.

<sup>47</sup>Bagger, Moen, and Vejlin (2017) provides a normative income taxation analysis in the context of a frictional labor market with two-sided heterogeneity.

<sup>48</sup>An obvious alternative would be to look for a flexible, say, piece-wise linear tax system as in Blundell and Shephard (2012).

Figure 2.8: A Pareto optimal tax reform

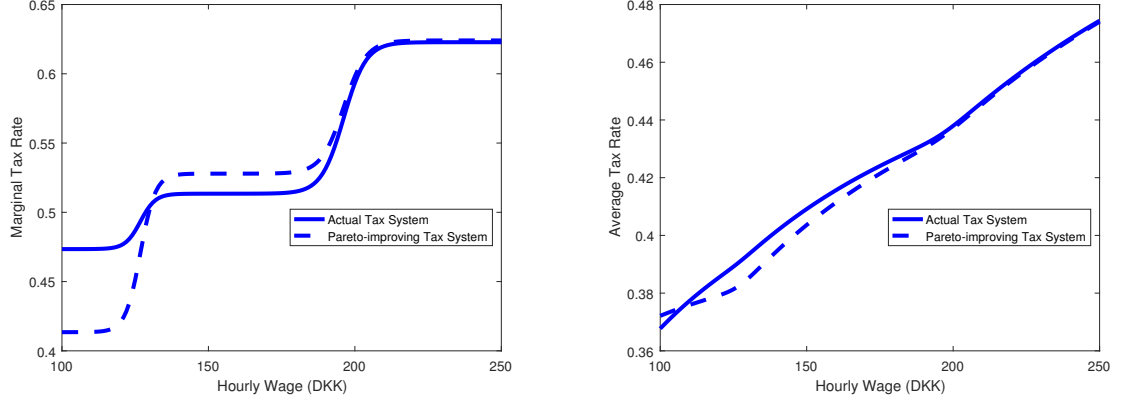


Table 2.11: Percentage changes between 1999-2003 regime and optimally reformed regime—Partial Equilibrium

Ability	$S_0$	$S_1$	$V$	$\theta$	$n^0$	$N2J$	$J2J$	$w^F$
All	0.95	-0.17	0.00	0.00	-1.37	1.96	-0.12	0.00
0.25	2.53	5.04	0.00	0.00	-5.44	7.47	0.91	0.00
0.5	0.12	-0.96	0.00	0.00	-0.79	0.87	-0.42	0.00
0.75	0.01	-0.25	0.00	0.00	-0.12	0.13	-0.12	0.00

The actual reforms tended to uniformly reduce the marginal tax rates.

To get a sense of how the social planner achieves this welfare gain Tables 2.11 and 2.12 calculate the percentage change in a number of key variables between the 1999-2003 regime and the optimally reformed regime, considering both the aggregate effects across the ability distribution, as well as showing effects for specific ability-types (the 25th, 50th and 75th percentiles in the ability distribution). The variables we consider are aggregate unemployed search  $S_0$ , aggregate employed search  $S_1$ , aggregate hiring intensity  $V$ , market tightness  $\theta$ , the nonemployment rate  $n^0$ , the aggregate non-employment to employment rate  $N2J$ , the aggregate job-to-job transition rate ( $J2J$ ), and the average percentage change in the wages offered by firms  $w^F$ . Table 2.11 shows the partial equilibrium effects, whereas Table 2.12 show the full equilibrium effects.

From Table 2.11 we see that in partial equilibrium, where firm behavior is fixed, the optimally reformed tax system increases unemployed search and therefore reduces the unemployment rate. This reduces the planner's expenditures on unemployment benefits. This effect of the optimally reformed income tax regime is especially pronounced at the bottom of the ability distribution—a group of workers that also faces high unemployment

Table 2.12: Percentage changes between 1999-2003 regime and optimally reformed regime—Full Equilibrium

Ability	$S_0$	$S_1$	$V$	$\theta$	$n^0$	$N2J$	$J2J$	$w^F$
All	2.20	-1.14	0.75	-0.29	-3.11	4.62	-0.89	0.10
0.25	3.54	4.09	-0.20	-2.61	-6.67	9.41	0.35	0.47
0.5	-0.05	-3.31	0.87	1.85	-0.75	0.37	-1.74	0.01
0.75	0.00	-0.24	0.57	0.66	-0.10	-0.18	-0.41	0.03

rates, and are likely to be located at the bottom of the wage distribution when in work. Table 2.11 also shows that, in partial equilibrium, employed job search and hence the job-to-job transition rate are reduced. This comes from the fact that the optimally reformed marginal income tax rates in the middle and towards the top of the wage distribution are increased, reducing the incentives to search for those jobs. Workers in the bottom quartile of the ability distribution actually increases their employed job search, because the reduced tax rates at the lower wages impact most of the wage ladder that these workers face.

Turning now to the full equilibrium responses tabulated in Table 2.12 we note that there is a qualitatively similar full equilibrium impact on  $S_0$ ,  $S_1$ ,  $n_0$ ,  $N2J$  and  $J2J$ . However, firm responses to the optimally reformed tax system involves an increase in hiring intensity, and an increase in the offered wages. This equilibrium response tend to magnify the partial equilibrium responses described above. Hence, the intuition underlying the optimally reformed income tax schedule remains that it moves especially low ability workers out of nonemployment, and higher up on the wage ladder.

**Conditioning taxes on ability.** Finally, we discuss the possibility of conditioning the tax schedule on individual ability. In the model, ability  $a$  is observable to both workers and firms, as well as the planner designing the Pareto optimal tax reform. It is therefore natural to consider the case where the planner can condition taxes on worker ability. This gives additional degrees of freedom to the social planner, and we discuss here the potential for further welfare gains through the introduction of an ability-conditional tax function,  $T(w; a)$ .

From the above discussion of the optimally reformed tax system we noticed that the social planner achieved the welfare gains primarily by targeting the behavior of low ability workers with high propensity for nonemployment. However, without being able to condition tax rates on worker-ability the targeting is less than perfect. Indeed, the optimally reformed tax system does not change the behavior of workers of median or higher ability.



We conjecture therefore that an ability-specific tax function would be able to substantially improve welfare relative to the one depicted in Figure 2.8. Most of these additional gains are likely to come through induced behavioral changes among workers above the 25th percentile of the ability distribution. We are currently working on  $a$ -specific tax functions  $T(w; a)$  in this Pareto-optimal tax reform exercise.

## 2.9 Conclusion

Income taxation reduces the return to job search effort, thus distorting workers' job search effort, and therefore the equilibrium allocation of heterogeneous workers across heterogeneous firms. Using a rich equilibrium frictional labor market model of search on-the-job, we have documented economically significant effects of income taxation on the allocation of labor. Specifically, we found that 1994, 1999, and 2004 Danish income tax reforms improved the equilibrium allocation of labor, reducing the steady state unemployment rate by about 6 percent, and increasing steady state labor income by about 2 percent, with workers of low ability gaining the most in terms of lower unemployment rates and higher labor incomes.

We also computed a mobility-based long-run elasticity of taxable labor income to be 0.12, with the labor income of lower ability workers being more responsive to income taxation than that of high ability workers. A decomposition of the elasticity reveals that the larger response of low ability workers is mostly along the extensive margin, whereas the smaller adjustments towards the top of the ability distribution are along the intensive margin and the wage policy margin. Finally, we have identified a Pareto improving labor income taxation reform that increases government revenue by 0.7 percent while leaving all workers weakly better off. Throughout the paper we have emphasized that the distinctions between partial and full equilibrium, income and substitution effects, as well as worker heterogeneity are quantitatively important.

This paper has dealt exclusively with positive aspects of labor income taxation in a framework with labor market frictions and two-sided heterogeneity where the allocation of labor is endogenous with respect to the tax system. The model set up here does not lend itself easily to normative analysis, but in a related paper Bagger et al. (2017) studies optimal taxation in a similar environment, within a model better suited for normative analysis.

## Appendix A

## Appendices

## A.1 Proofs and derivations

### A.1.1 Law of motions for $n_t^0(a)$ and $n_t^1(w|a)$

Recall from the main text that  $n_t^0(a)$  is the time- $t$  measure of unemployed workers in the household, while  $n_t^1(w|a)$  is the measure of workers employed with wage  $w$ . Define

$$N_t(a) \equiv n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} n_t^1(w|a)dw \quad (\text{A.1})$$

as the time- $t$  measure of the household population.

Now, consider a small interval of time of length  $dt$  from  $t$  to  $t + dt$ . The change in the type- $a$  household-specific unemployment rate from  $t$  to  $t + dt$  is given by the difference in gross worker flows in and out of unemployment. Indeed,

$$n_{t+dt}^0(a) - n_t^0(a) = \delta(a)dt[N_t(a) - n_t^0(a)] - \lambda dt [\mu + s_t^0(a)]\bar{F}(\phi(a)|a)n_t^0(a). \quad (\text{A.2})$$

There are  $N_t(a) - n_t^0(a)$  employed workers at time  $t$ , and within a small interval of time of length  $dt$ , a share  $\delta(a)dt$  of them have their job destroyed, generating a gross flow  $\delta(a)dt[N_t(a) - n_t^0(a)]$  into unemployment. There are  $n_t^0(a)$  unemployed workers in the type- $a$  household at time  $t$ , and between  $t$  and  $t + dt$ , a share  $\lambda dt [\mu + s_t^0(a)]\bar{F}(\phi(a)|a)$  of them obtain an acceptable job offer, generating a gross flow  $\lambda dt [\mu + s_t^0(a)]\bar{F}(\phi(a)|a)n_t^0(a)$  out of unemployment.

In equilibrium,  $\bar{F}(\phi(a)|a) = 1$  as no firm would optimally set the offered type- $a$  worker wage below the type- $a$  household's reservation wage. Imposing  $\bar{F}(\phi(a)|a) = 1$  and dividing through by  $dt > 0$  and taking the limit as  $dt \rightarrow 0$  yields the law of motion (2.6) stated in the main text.

The law of motion for  $n_t^1(w|a)$  derives from a similar logic. Consider the change in the measure of workers employed at a job paying  $w$  or less during a small interval of time of length  $d$  from  $t$  to  $t + dt$ . That is, let  $N_t^1(w|a) = \int_{\phi(a)}^w n_t^1(w|a)dw$ , and consider

$$\begin{aligned} N_{t+dt}^1(w|a) - N_t^1(w|a) = & \\ \lambda dt [\mu + s_t^0(a)][F(w|a) - F(\phi(a)|a)]n_t^0(a) + \lambda dt \mu [F(w|a) - F(\phi(a)|a)][N_t(a) - n_t^0(a) - N_t^1(w|a)] & \\ - [\delta(a) + \lambda \mu F(\phi(a)|a)]dt N_t^1(w|a) - \lambda dt \mu \bar{F}(w|a)N_t^1(w|a) - \bar{F}(w|a) \int_{\phi(a)}^w \lambda dt s_t^1(x|a)n_t^1(x|a)dx. & \end{aligned} \quad (\text{A.3})$$

The first term stems from the gross inflow of unemployed workers who find an acceptable job paying  $w$  or less. The second term stems from the inflow of workers currently employed

at a wage greater than  $w$  (there are  $N_t(a) - n_t^0(a) - N_t^1(w|a)$  such workers in a type- $a$  household), but who are reallocated to an acceptable job paying  $w$  or less. The sum of the first two terms constitutes total inflow into the group of employed type- $a$  household members earning a wage  $w$  or below. The third term reflects the gross worker outflow made up of employed workers earning a wage  $w$  or below who move into unemployment. They may do so due to a job destruction shock, or a reallocation shock involving an unacceptable job offer, an event that occurs with probability  $\lambda\mu F(\phi(a)|a)dt$  in a short interval of length  $dt$ . The fourth term comprises gross worker outflow from workers earning a wage  $w$  or less who are reallocated to jobs paying in excess of  $w$ . Finally, the fifth term reflects gross outflow of workers who, through on-the-job search, located a job paying in excess of  $w$ . The sum of the last three terms constitutes total outflow from the group of employed type- $a$  household members earning a wage  $w$  or below.

Rearranging and collecting terms in (A.3), imposing the equilibrium restriction  $F(\phi(a)|a) = 0$ , dividing through by  $dt > 0$  and taking the limit as  $dt \rightarrow 0$  yields

$$\begin{aligned} \dot{N}_t^1(w|a) = & [\lambda\mu + \lambda s_t^0(a)]F(w|a)n_t^0(a) + \lambda\mu F(w|a)[N_t(a) - n_t^0(a) - N_t^1(w|a)] \\ & - \delta(a)N_t^1(w|a) - \lambda\mu\bar{F}(w|a)N_t^1(w|a) - \bar{F}(w|a) \int_{\phi(a)}^w \lambda s_t^1(x|a)n_t^1(x|a)dx. \end{aligned} \quad (\text{A.4})$$

Finally, taking the derivative with respect to  $w$  and a bit of algebra yields the law of motion (2.7) stated in the main text.

### A.1.2 Steady state labor allocation

The size of a household is  $N_t(a) = n_t^0(a) + \int_{\psi(a)}^{\bar{w}} n_t^1(w|a)$ . Hence,  $n_t^0(a)/N_t(a)$  is the type- $a$  household time- $t$  unemployment rate. The PDF of employed workers across wage rates is  $g_t(w|a) = n_t^1(w|a)/[1 - n_t^0(a)]$ , with associated CDF  $G_t(w|a) = \int_{\phi(a)}^{\bar{w}} g_t(w|a)dw = N_t^1(w|a)/[1 - n_t^0(a)]$ . The law of motions for  $n_t^0(a)$  and  $n_t^1(w|a)$  are discussed above.  $G_t(w|a)$  is the steady state employment-weighted wage distribution, which reflects the allocation of labor across firms. The law of motion for  $G_t(w|a)$  is straightforward to obtain from (A.4). Indeed,

$$\begin{aligned} \dot{G}_t(w|a) = & [\lambda\mu + \lambda s_t^0(a)]F(w|a)\frac{n_t^0(a)}{1 - n_t^0(a)} + \lambda\mu F(w|a)\bar{G}_t(w|a) \\ & - \delta(a)G_t(w|a) - \lambda\mu\bar{F}(w|a)G_t(w|a) - \bar{F}(w|a) \int_{\phi(a)}^w \lambda s_t^1(x|a)g_t(x|a)dx. \end{aligned} \quad (\text{A.5})$$

In steady state  $\dot{n}_t^0(a) = 0$ , and  $s_t^0(a) = s^0(a)$ , and  $\dot{n}_t^1(w|a) = 0$  and  $s_t^1(w|a) = s^1(w|a)$  for all  $w \in [\psi(a), \bar{w}(a)]$ , in which case (A.2) implies that the worker-type conditional

steady state unemployment rate,  $n^0(a)$ , is

$$n^0(a) = \frac{\delta(a)}{\delta(a) + \lambda\mu + \lambda s^0(a)}. \quad (\text{A.6})$$

Furthermore, in steady state,  $\dot{G}_t(w|a) = 0$ , in which case (A.5) simplifies as

$$\int_{\phi(a)}^w \lambda s^1(x|a) g(x|a) dx = [\delta(a) + \lambda\mu] \frac{F(w|a) - G(w|a)}{\bar{F}(w|a)}, \quad (\text{A.7})$$

where we have used that, according to (A.6),  $n^0(a)/[1 - n^0(a)] = \delta(a)/[\lambda\mu + \lambda s^0(a)]$ . Next, take the derivative of (A.7) with respect to  $w$  to obtain

$$\lambda s^1(w|a) g(w|a) = [\delta(a) + \lambda\mu] \frac{f(w|a) \bar{G}(w|a) - g(w|a) \bar{F}(w|a)}{\bar{F}(w|a)^2}, \quad (\text{A.8})$$

which, upon multiplying through by  $\bar{F}(w|a)/\bar{G}(w|a)$ , yields

$$\lambda s^1(w|a) \bar{F}(w|a) \frac{g(w|a)}{\bar{G}(w|a)} = [\delta(a) + \lambda\mu] \frac{f(w|a)}{\bar{F}(w|a)} - [\delta(a) + \lambda\mu] \frac{g(w|a)}{\bar{G}(w|a)}. \quad (\text{A.9})$$

Rearrange to get an expression for the hazard function of  $G(w|a)$ ,

$$\frac{g(w|a)}{\bar{G}(w|a)} = \left[ \frac{\delta(a) + \lambda\mu}{\delta(a) + \lambda\mu + \lambda s^1(w|a) \bar{F}(w|a)} \right] \frac{f(w|a)}{\bar{F}(w|a)}. \quad (\text{A.10})$$

It follows that the steady state labor allocation,  $G(w|a)$ , is given by

$$G(w|a) = 1 - \exp \left\langle - \int_{\phi(a)}^w \left[ \frac{\delta(a) + \lambda\mu}{\delta(a) + \lambda\mu + \lambda s^1(x|a) \bar{F}(x|a)} \right] \frac{f(x|a)}{\bar{F}(x|a)} dx \right\rangle. \quad (\text{A.11})$$

As mentioned in the main text, this expression is useful for the numerical solution of the model because it allow us to compute the allocation of labor from any search effort function  $s^1(w|a)$  and wage offer distribution  $F(x|a)$  without having to (numerically) solve a system of balanced flow equations.

**The special case of wage-independent search effort,**  $s^1(w|a) = s^1$ . While useful, the derivation of, and the resulting expression for, the employment-weighted wage distribution  $G(w|a)$  is somewhat unconventional. Consider the special case where on-the-job search effort is independent of the wage. This exogenous search effort specification is common in the job search literature. Under the restriction that  $s^1(w|a) = s^1(a)$ , the

integrated hazard function of the employment-weighted wage distribution is given by

$$\begin{aligned}
\int_{\phi(a)}^w \frac{g(x|a)}{\bar{G}(x|a)} dx &= \int_{\phi(a)}^w \left[ \frac{\delta(a) + \lambda\mu}{\delta(a) + \lambda\mu + \lambda s^1(a) \bar{F}(x|a)} \right] \frac{f(x|a)}{\bar{F}(x|a)} dx \\
&= \int_{\phi(a)}^w [1 + \kappa^1(a) \bar{F}(x|a)]^{-1} \frac{f(x|a)}{\bar{F}(x|a)} dx \\
&= \int_{\bar{F}(w|a)}^1 [1 + \kappa^1(a) z]^{-1} \frac{1}{z} dz \\
&= \int_{\bar{F}(w|a)}^1 \left[ \frac{1}{z} - \frac{\kappa^1(a)}{1 + \kappa^1(a) z} \right]^{-1} dz. \tag{A.12}
\end{aligned}$$

where  $\kappa^1(a) \equiv \frac{\lambda s^1(a)}{\delta(a) + \lambda\mu}$ . Integrating (A.12) yields

$$\begin{aligned}
\int_{\phi(a)}^w \frac{g(x|a)}{\bar{G}(x|a)} dx &= \left|_{\bar{F}(w|a)}^1 \ln \left( \frac{z}{1 + \kappa^1(a) z} \right) \right. \\
&= \ln \left( \frac{1}{1 + \kappa^1(a)} \right) - \ln \left( \frac{\bar{F}(w|a)}{1 + \kappa^1(a) \bar{F}(w|a)} \right) \\
&= \ln \left( \frac{1 + \kappa^1(a) \bar{F}(w|a)}{[1 + \kappa^1(a)] \bar{F}(w|a)} \right), \tag{A.13}
\end{aligned}$$

Hence, using (A.11) on (A.13), we obtain

$$G(w|a) = 1 - \exp \left\langle -\ln \left( \frac{1 + \kappa^1(a) \bar{F}(w|a)}{[1 + \kappa^1(a)] \bar{F}(w|a)} \right) \right\rangle = 1 - \frac{[1 + \kappa^1(a)] \bar{F}(w|a)}{1 + \kappa^1(a) \bar{F}(w|a)} = \frac{F(w|a)}{1 + \kappa^1(a) \bar{F}(w|a)}. \tag{A.14}$$

We note that (A.14) is the “standard” expression for the employment-weighted distribution of wages that appears in e.g. Burdett and Mortensen (1998) and Bontemps et al. (2000).

### A.1.3 Applying the Maximum Principle

The current value Hamiltonian associated with the household problem is given as

$$\begin{aligned}
\mathcal{H}_t(a) &= \psi(c_t(a), s_t^0(a)) n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} \psi(c_t(a), s_t^1(w|a)) n_t^1(w|a) dw \\
&\quad + e^{\rho t} \xi_t^0(a) \left\langle \delta(a) N_t(a) - [\delta(a) + \lambda\mu + \lambda s_t^0(a)] n_t^0(a) \right\rangle \\
&\quad + \int_{\phi(a)}^{\bar{w}(a)} e^{\rho t} \xi_t^1(x|a) \left\langle \left[ \lambda\mu N_t(a) + \lambda s_t^0(a) n_t^0(a) + \int_{\phi(a)}^x \lambda s_t^1(z|a) n_t^1(z|a) dz \right] f(x|a) \right. \\
&\quad \quad \left. - \left[ \delta(a) + \lambda\mu + \lambda s_t^1(x|a) \bar{F}(x|a) \right] n_t^1(x|a) \right\rangle dx, \tag{A.15}
\end{aligned}$$

with  $c_t(a)$  given by (2.4), restated here for later reference,

$$c_t(a) = \frac{1}{N_t(a)} \left( [b - T(b)]n_t^0(a) + \int_{\phi(a)}^{\bar{w}(a)} [w - T(w)]n_t^1(w|a)dw \right) + \bar{b} \quad (\text{A.16})$$

and  $N_t(a)$  given by (A.1).

By the Maximum Principle (Acemoglu, 2009, Theorem 7.13, p. 254), the optimal controls satisfies the following necessary conditions. First, with respect to employed job search effort  $s_t^1(w|a)$  at any wage level  $w \in [\phi(a), \bar{w}(a)]$ , optimality requires

$$\frac{\partial \mathcal{H}_t(a)}{\partial s_t^1(w|a)} = 0 \quad \text{for all } t \in [0, \infty) \text{ and for all } w \in [\phi(a), \bar{w}(a)], \quad (\text{A.17})$$

$$\frac{\partial \mathcal{H}_t(a)}{\partial n_t^1(w|a)} = -e^{\rho t} \dot{\xi}_t^1(w|a) \quad \text{for all } t \in [0, \infty) \text{ and for all } w \in [\phi(a), \bar{w}(a)], \quad (\text{A.18})$$

where  $\mathcal{H}_t(a)$  is given by (A.15) and  $\dot{\xi}_t^1(w|a) \equiv \partial \xi_t^1(w|a) / \partial t$ . Of course, there is set of analogous necessary optimality conditions pertaining to unemployed search effort  $s_t^0(a)$ . Indeed,

$$\frac{\partial \mathcal{H}_t(a)}{\partial s_t^0(a)} = 0 \quad \text{for all } t \in [0, \infty), \quad (\text{A.19})$$

$$\frac{\partial \mathcal{H}_t(a)}{\partial n_t^0(a)} = -e^{\rho t} \dot{\xi}_t^0(a) \quad \text{for all } t \in [0, \infty). \quad (\text{A.20})$$

We next solve the system of equation made up (A.19), (A.20), (A.17), and (A.18), with  $\mathcal{H}_t(a)$  given by (A.15).

**Optimal steady state job search effort,  $s_t^1(w|a)$  and  $s_t^0(a)$ .** Consider first employed job search effort,  $s_t^1(w|a)$ . Using (A.15), equating the derivative of  $\mathcal{H}_t(a)$  with respect to  $s_t^1(w|a)$ , for some  $w \in [\phi(a), \bar{w}(a)]$ , to zero, and rearranging, yields

$$e^{-\rho t} \zeta'(s_t^1(w|a)) = \lambda \int_w^{\bar{w}(a)} [\xi_t^1(x|a) - \xi_t^1(w|a)] f(x|a) dx, \quad (\text{A.21})$$

for any  $w \in [\phi(a), \bar{w}(a)]$ , and where  $\zeta'(\cdot)$  represents the marginal disutility of job search effort, see (2.2). The necessary optimality conditions (A.21) has a straightforward interpretation that is spelled out in the main text. For future reference, notice that integrating (A.21) by parts implies

$$e^{-\rho t} \zeta'(s_t^1(w|a)) = \lambda \int_w^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_t^1(x|a) \bar{F}(x|a) dx. \quad (\text{A.22})$$

Next, consider unemployed job search effort  $s_t^0(a)$ . Derivations analogous to those that lead to (A.21) implies that

$$e^{-\rho t} \zeta'(s_t^0(a)) = \lambda \int_{\phi(a)}^{\bar{w}(a)} [\xi_t^1(x|a) - \xi_t^0(a)] f(x|a) dx, \quad (\text{A.23})$$

and integration by parts yields

$$e^{-\rho t} \zeta'(s_t^0(a)) = \lambda \int_{\phi(a)}^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_t^1(x|a) \bar{F}(x|a) dx. \quad (\text{A.24})$$

**The steady state costate  $\xi_t^1(x|a)$ .** Consider first the costate variable  $\xi_t^1(x|a)$ , the multiplier on the law of motion for the control  $n_t^1(w|a)$ , the time- $t$  measure of workers employed at wage  $w$  in the type- $a$  household. Utilizing (A.18) and equating the derivative of  $\mathcal{H}_t(a)$  with respect to the control  $n_t^1(w|a)$  to  $-e^{\rho t} \dot{\xi}_t^1(w|a)$  yields

$$\begin{aligned} -\dot{\xi}_t^1(w|a) = e^{-\rho t} & \left[ u(c_t(a)) - \zeta(s_t^1(w|a)) + u'(c_t(a))[w - T(w) - c_t(a)] \right] \\ & + \delta(a)[\xi_t^0(a) - \xi_t^1(w|a)] + \lambda \mu \int_{\phi(a)}^{\bar{w}(a)} [\xi_t^1(x|a) - \xi_t^1(w|a)] f(x|a) dx \\ & + \lambda s_t^1(w|a) \int_w^{\bar{w}(a)} [\xi_t^1(x|a) - \xi_t^1(w|a)] f(x|a) dx, \quad (\text{A.25}) \end{aligned}$$

for any  $w \in [\phi(a), \bar{w}(a)]$ . In deriving (A.25) we have used that  $\partial N_t(a)/\partial n_t^1(w|a) = 1$ , and that, by (A.16),

$$\frac{\partial c_t(a)}{\partial n_t^1(w|a)} = \frac{w - T(w) - c_t(a)}{N_t(a)}. \quad (\text{A.26})$$

The first term on the right-hand side of (A.25) is the change in the household's flow-payoff if a worker earning a wage  $w$  is added to the household at time  $t$ . Such an additional worker increases the time- $t$  utility flow in the household by  $u(c_t(a)) - \zeta(s_t^1(w|a))$  and, in addition, contributes net-of-tax revenue in the amount of  $w - T(w) - c_t(a)$  to be distributed within the household. The household values this additional revenue at the marginal utility of consumption,  $u'(c_t(a))$ . The two last terms on the right-hand side of (A.25) represent the change in the value of the stock of workers earning a wage  $w$ , denominated in present values, stemming from the additional worker. This change is given by the expected capital gains from job destruction, reallocation and job search at the dictated search effort,  $\lambda s_t^1(w|a)$  that the additional worker is subject to. At the optimal search policy, the change in the current flow-payoff plus the change to the value of the household's stock of workers employed at wage  $w$ ,  $n_t^1(w|a)$ , balance the depreciation of the stock, i.e.  $-\dot{\xi}_t^1(w|a)$ .



Substituting (A.21) into (A.25) and rearranging allow us to write

$$-\dot{\xi}_t^1(w|a) = e^{-\rho t} \Phi_t^1(w|a) - [\delta(a) + \lambda\mu] \xi_t^1(w|a), \quad (\text{A.27})$$

where  $\Phi_t^1(w|a)$  is defined as

$$\begin{aligned} \Phi_t^1(w|a) \equiv & u(c_t(a)) - \zeta(s_t^1(w|a)) + u'(c_t(a))[w - T(w) - c_t(a)] \\ & + \delta(a)e^{\rho t} \xi_t^0(a) + \lambda\mu \int_{\phi(a)}^{\bar{w}(a)} e^{\rho t} \xi_t^1(x|a) f(x|a) dx + \zeta'(s_t^1(w|a)) s_t^1(w|a). \end{aligned} \quad (\text{A.28})$$

Solving the non-homogenous first-order linear ordinary differential equation (ODE) (A.27) for  $\xi_t^1(w|a)$  yields

$$\xi_t^1(w|a) = e^{[\delta(a)+\lambda\mu]t} \xi_0^1(w|a) - e^{[\delta(a)+\lambda\mu]t} \int_0^t e^{-[\rho+\delta(a)+\lambda\mu]x} \Phi_x^1(w|a) dx. \quad (\text{A.29})$$

Consider now an economy in steady state, i.e. consider the situation where  $\dot{n}_t^1(w|a) = 0$  for all  $w \in [\phi(a), \bar{w}(a)]$ , as well as  $\dot{n}_t^0(a) = 0$ , such that  $n_t^1(w|a) = n_{t'}^1(w|a) = n^1(w|a)$  for all  $w \in [\phi(a), \bar{w}(a)]$ ,  $n_t^0(a) = n_{t'}^0(a) = n^0(a)$ , and where  $s_t^1(w|a) = s_{t'}^1(w|a) = s^1(w|a)$  and  $s_t^0(a) = s_{t'}^0(a) = s^0(a)$  for any two dates  $t$  and  $t'$ . The steady state version (A.28) implies that  $\Phi_t^1(w|a) = \Phi_{t'}^1(w|a) = \Phi^1(w|a)$  for any two dates  $t$  and  $t'$ . Moreover, in steady state, the *current value* costate variables are equalized, i.e.  $\xi_0^1(w|a) = e^{\rho t} \xi_t^1(w|a)$  for all  $w \in [\phi(a), \bar{w}(a)]$ . Put differently, the *present value* costate variables differ only across time due to discounting.<sup>1</sup> The steady state version of (A.29) therefore implies that

$$\xi_t^0(w|a) = \frac{e^{[\delta(a)+\lambda\mu]t}}{1 - e^{[\rho+\delta(a)+\lambda\mu]t}} \left[ 1 - e^{-[\rho+\delta(a)+\lambda\mu]t} \right] \frac{\Phi^1(w|a)}{\rho + \delta(a) + \lambda\mu}. \quad (\text{A.30})$$

The transversality condition for the costate  $\xi_t^1(w|a)$  holds. Indeed, since

$$\lim_{t \rightarrow \infty} \frac{e^{[\delta(a)+\lambda\mu]t}}{1 - e^{[\rho+\delta(a)+\lambda\mu]t}} = -\frac{\delta(a) + \lambda\mu}{\rho + \delta(a) + \lambda\mu} \lim_{t \rightarrow \infty} e^{-\rho t} = 0, \quad (\text{A.31})$$

we have

$$\lim_{t \rightarrow \infty} \xi_t^1(w|a) = 0. \quad (\text{A.32})$$

We note that the steady state identity  $\xi_0^1(w|a) = e^{\rho t} \xi_t^1(w|a)$  and (A.30) implies that the household's steady state shadow valuation of an additional worker employed with wage

---

<sup>1</sup>Recall from the main text that  $\xi_t^1(w|a)$  is the present value shadow value of adding an additional employed worker earning a wage  $w$  to the household at time  $t$ . In steady state, this differs from the present value shadow value of adding an additional employed worker at wage rate  $w$  to the household at time 0,  $\xi_0^1(w|a)$ , only by discounting. Hence,  $\xi_0^1(w|a) = e^{\rho t} \xi_t^1(w|a)$ .

rate  $w$  is

$$[\rho + \delta(a) + \lambda\mu]\xi_0^1(w|a) = \Phi^1(w|a), \quad (\text{A.33})$$

which, upon substituting the steady state version of (A.28) into (A.33), implies that

$$\begin{aligned} \rho\xi_0^1(w|a) &= u(c(a)) - \zeta(s^1(w|a)) + u'(c(a))[w - T(w) - c(a)] \\ &+ \delta(a)[\xi_0^0(a) - \xi_0^1(w|a)] + \lambda\mu \int_{\phi(a)}^{\bar{w}(a)} [\xi_0^1(x|a) - \xi_0^1(w|a)]f(x|a)dx + \zeta'(s^1(w|a))s^1(w|a), \end{aligned} \quad (\text{A.34})$$

where  $c(a)$  is steady state household consumption, see (A.16). Substituting (A.22) back into (A.34), using again the steady state identity  $\xi_0^1(w|a) = e^{\rho t}\xi_t^1(w|a)$ , and integrating the remaining integral by parts yield

$$\begin{aligned} \rho\xi_0^1(w|a) &= u(c(a)) - \zeta(s^1(w|a)) + u'(c(a))[w - T(w) - c(a)] \\ &+ \delta(a)[\xi_0^0(a) - \xi_0^1(w|a)] + \lambda\mu \int_{\phi(a)}^{\bar{w}(a)} \frac{\partial}{\partial x}\xi_0^1(x|a)\bar{F}(x|a)dx + \lambda s^1(w|a) \int_w^{\bar{w}(a)} \frac{\partial}{\partial x}\xi_0^1(x|a)\bar{F}(x|a)dx. \end{aligned} \quad (\text{A.35})$$

By the Envelope Theorem, the derivative of (A.35) with respect to  $w$  implies that

$$\frac{\partial}{\partial w}\xi_0^1(w|a) = \frac{u'(c(a))[1 - T'(w)]}{\rho + \delta(a) + \lambda\mu + \lambda s^1(w|a)\bar{F}(w|a)}. \quad (\text{A.36})$$

From (A.36), we note that  $\frac{\partial}{\partial w}\xi_0^1(w|a) > 0$ , and hence, that  $\xi_0^1(w|a)$  is strictly increasing in  $w$ . This implies that the household dictates that employed workers accept an alternative job offer if and only if it pays a higher wage than the current job.

Substituting (A.36) back into (A.35), allow us to express the steady state costate  $\xi_0^1(w|a)$  as

$$\begin{aligned} \rho\xi_0^1(w|a) &= u(c(a)) - \zeta(s^1(w|a)) + u'(c(a))[w - T(w) - c(a)] \\ &+ \delta(a)[\xi_0^0(a) - \xi_0^1(w|a)] + \lambda\mu \int_{\phi(a)}^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s^1(x|a)\bar{F}(x|a)}dx \\ &+ \lambda s^1(w|a) \int_w^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s^1(x|a)\bar{F}(x|a)}dx. \end{aligned} \quad (\text{A.37})$$

The interpretation of (A.37) is described in the main text.

Finally, substituting (A.36) into (A.22) yields a useful characterization of the optimal

steady state search effort among employed workers,

$$\zeta'(s^1(w|a)) = \lambda \int_w^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s^1(x|a)\bar{F}(x|a)} dx. \quad (\text{A.38})$$

**The steady state costate  $\xi_t^0(a)$ .** The characterization of the costate  $\xi_t^0(a)$  is analogous to that of  $\xi_t^1(w|a)$  detailed above, but is nonetheless included here for completeness. First, equate the derivative of  $\mathcal{H}_t(a)$  with respect to  $n_t^0(a)$  to  $-e^{\rho t}\dot{\xi}_t^0(a)$ , to obtain

$$\begin{aligned} -\dot{\xi}_t^0(a) = e^{-\rho t} & \left[ u(c_t(a)) - \zeta(s_t^0(a)) + u'(c_t(a))[b - T(b) - c_t(a)] \right] \\ & + [\lambda\mu + \lambda s_t^0(a)] \int_{\phi(a)}^{\bar{w}(a)} [\xi_t^1(x|a) - \xi_t^0(a)] f(x|a) dx. \end{aligned} \quad (\text{A.39})$$

Equation (A.39) has the same interpretation as (A.25).

Substituting (A.23) into (A.39) and rearranging yields

$$-\dot{\xi}_t^0(a) = e^{-\rho t} \Phi_t^0(a) - \lambda\mu \xi_t^0(a), \quad (\text{A.40})$$

where  $\Phi_t^0(a)$  is defined as

$$\Phi_t^0(a) \equiv u(c_t(a)) - \zeta(s_t^0(a)) + u'(c_t(a))[b - T(b) - c_t(a)] + \int_{\phi(a)}^{\bar{w}(a)} e^{\rho t} \xi_t^1(x|a) f(x|a) dx + \zeta'(s_t^0(a)) s_t^0(a). \quad (\text{A.41})$$

The solution to the non-homogenous first-order linear ODE (A.40) can be expressed as

$$\xi_t^0(a) = e^{\lambda\mu t} \xi_0^0(a) - e^{\lambda\mu t} \int_0^t e^{-[\rho + \lambda\mu]x} \Phi_x^0(a) dx. \quad (\text{A.42})$$

Imposing steady state, where  $\Phi_t^0(a) = \Phi^0(a)$  and  $\xi_0^0(a) = e^{\rho t} \xi_t^0(a)$ , on (A.42) implies

$$\xi_t^0(a) = \frac{e^{\lambda\mu t}}{1 - e^{[\rho + \lambda\mu]t}} \left[ 1 - e^{-[\rho + \lambda\mu]t} \right] \frac{\Phi^0(a)}{\rho + \lambda\mu}. \quad (\text{A.43})$$

It is straightforward to verify that the transversality condition  $\lim_{t \rightarrow \infty} \xi_t^0(a) = 0$  holds, and that the household's steady state shadow valuation of an additional unemployed worker,  $\xi_t^0(a)$  is

$$[\rho + \lambda\mu] \xi_0^0(a) = \Phi^0(a). \quad (\text{A.44})$$

Substituting the steady state version of (A.41) into (A.44) yields

$$\rho\xi_0^0(a) = u(c(a)) - \zeta(s^0(a)) + u'(c(a))[b - T(b) - c(a)] + \lambda\mu \int_{\phi(a)}^{\bar{w}(a)} [\xi_0^1(x|a) - \xi_0^0(a)] f(x|a) dx + \zeta'(s^0(a))s^0(a), \quad (\text{A.45})$$

where  $c(a)$  is steady state household consumption, see (A.16). Substitute (A.24) into (A.45) and integrate the integral by parts to obtain

$$\rho\xi_0^0(a) = u(c(a)) - \zeta(s^0(a)) + u'(c(a))[b - T(b) - c(a)] + \lambda[\mu + s^0(a)] \int_{\phi(a)}^{\bar{w}(a)} \frac{\partial}{\partial x} \xi_0^1(x|a) \bar{F}(x|a) dx. \quad (\text{A.46})$$

The derivative  $\frac{\partial}{\partial x} \xi_0^1(x|a)$  is given by (A.36), and hence, we may write

$$\begin{aligned} \rho\xi_0^0(a) = & u(c(a)) - \zeta(s^0(a)) + u'(c(a))[b - T(b) - c(a)] \\ & + \lambda[\mu + s^0(a)] \int_{\phi(a)}^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s^1(x|a)\bar{F}(x|a)} dx. \end{aligned} \quad (\text{A.47})$$

Finally, substituting (A.36) into (A.24) yields,

$$\zeta'(s^0(a)) = \lambda \int_{\phi(a)}^{\bar{w}(a)} \frac{u'(c(a))[1 - T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s^1(x|a)\bar{F}(x|a)} dx. \quad (\text{A.48})$$

**The steady state reservation wage,  $\phi(a)$ .** The steady state reservation wage  $\phi(a)$  solves

$$\xi_0^1(\phi(a)|a) = \xi_0^0(a). \quad (\text{A.49})$$

In other words, it is the wage rate  $\phi(a)$  at which the household's shadow valuation of employment coincides with the shadow valuation of unemployment. From (A.37), (A.47), (A.38) and (A.48) it is evident that

$$\phi(a) = b. \quad (\text{A.50})$$

and that unemployed steady state search effort is identical to employed steady state search effort in the lowest paying firm with  $w = b$ ,

$$s^0(a) = s^1(b|a). \quad (\text{A.51})$$

#### A.1.4 Proof of Lemma 1

*Proof.* Consider property (i) in lemma 1. From (2.13),  $s(w|a) = [\zeta']^{-1} \left( \lambda \int_w^{\bar{w}(a)} \frac{u'(c(a))[1-T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s(x|a)\bar{F}(x|a)} dx \right) + \underline{s}$ . Since  $\zeta' : [0, \infty) \mapsto [0, \infty)$  we have  $[\zeta']^{-1} : [0, \infty) \mapsto [0, \infty)$ . Hence,  $[\zeta']^{-1}(\cdot) \geq 0$ , and therefore,  $s(w|a) \geq \underline{s}$ . Consider now property (ii) in lemma 1. The derivative of  $[\zeta']^{-1}(\cdot)$  is  $1/\zeta''([\zeta']^{-1}(\cdot))$ . Since  $\zeta''(\cdot) > 0$ ,  $[\zeta']^{-1} : [0, \infty) \mapsto [0, \infty)$  is a strictly increasing function. It follows that  $[\zeta']^{-1}(0) = 0$ , and hence,  $s(\bar{w}(a)|a) = 0$ . Finally, consider property (iii) in lemma 1.  $[\zeta']^{-1}(\cdot)$  is a strictly increasing function. Take  $w' < w''$ . Then,  $\int_{w'}^{\bar{w}(a)} \frac{u'(c(a))[1-T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s(x|a)\bar{F}(x|a)} dx > \int_{w''}^{\bar{w}(a)} \frac{u'(c(a))[1-T'(x)]\bar{F}(x|a)}{\rho + \delta(a) + \lambda\mu + \lambda s(x|a)\bar{F}(x|a)} dx$ . It follows that  $s(w'|a) > s(w''|a)$ .<sup>2</sup>  $\square$

#### A.1.5 Proof of Lemma 2

*Proof.* Take  $w' < w''$  both in  $[b, \bar{w}(a)]$ . Lemma 1 established that  $s(w'|a) > s(w''|a)$ , from which it follows that  $s(w'|a)\bar{F}(w'|a) > s(w''|a)\bar{F}(w''|a)$ . Finally,  $\int_b^{w'} s(x|a)dG(x|a) < \int_b^{w''} s(x|a)dG(x|a)$ . Lemma 2 now follows from (2.21) and (2.22).  $\square$

#### A.1.6 Proof of Proposition 1

*Proof.* Let  $p' < p''$  such that  $y(a, p') < y(a, p'')$ , and let  $w' = w(a|p')$  and  $w'' = w(a|p'')$ . Then,  $\tilde{\pi}(p''|w'', a) = [y(a, p'') - w'']\tilde{\ell}(a|w'') > [y(a, p'') - w']\tilde{\ell}(a|w') > [y(a, p') - w']\tilde{\ell}(a|w') = \tilde{\pi}(p'|w', a) > [y(a, p') - w'']\tilde{\ell}(a|w'')$ . This implies that  $[y(a, p'') - y(a, p')]\tilde{\ell}(a|w'') > [y(a, p'') - y(a, p')]\tilde{\ell}(a|w')$ , which implies  $\tilde{\ell}(a|w'') > \tilde{\ell}(a|w')$ . Lemma 2 established that  $\tilde{\ell}(a|w)$  is strictly increasing in  $w$ , which implies that  $w' < w''$ .  $\square$

#### A.1.7 Proof of Proposition 2

*Proof.* Fix  $a \in [0, 1]$ . The first order condition (FOC) for a type- $p$  firm's hiring intensity problem (2.23) is  $\tilde{\pi}(a|p)h(a)/V(a) = d'(v)$ . The left-hand side of the FOC is independent of  $v$ . Strict convexity of  $d(\cdot)$  implies that the right-hand side of the FOC is strictly increasing in  $v$ . Hence, the FOC has a unique solution  $v = v(a|p)$ . Strict convexity of  $d(\cdot)$  also ensures the second order condition is satisfied. The proof of Proposition 1 established that  $\tilde{\pi}(p|a)$  is strictly increasing in  $p$ . Strict convexity of  $d(\cdot)$  implies that  $v(a|p)$  is strictly increasing in  $p$ .  $\square$

<sup>2</sup>Alternatively, since  $F(w|a)$  is differentiable, direct differentiation of (2.13) yields the desired result.

### A.1.8 Proof of Proposition 3

*Proof.* Taking the derivative of (2.23), applying the Envelope Theorem, yields

$$\frac{\partial \pi(p)}{\partial p} = (1 - \tau) \int_0^1 \left[ \frac{v(a|p)}{V(a)} \right] \frac{\partial \tilde{\pi}(p|a)}{\partial p} dH(a).$$

The proof of Proposition 1 established that  $\tilde{\pi}(p|a)$  is strictly increasing in  $p$  for every  $a \in [0, 1]$ . Proposition 2 established that  $v(a|p)$  is strictly increasing in  $p$  for every  $a \in [0, 1]$ . The product of strictly increasing strictly positive functions yields a strictly increasing function. Integrating a set of strictly increasing functions yields a strictly increasing function. Proposition 3 follows.  $\square$

## A.2 Details on the data

### A.2.1 The merging procedure

Merging data sources by person ID is straightforward. However, a few complications arise when merging data sources by firm IDs because different public registers use different firm identifiers. Data clean-up, manipulation, merging and selection is carried out in a series of SAS-programs. Before we move on to describing the merging procedure, we provide a brief description of the person and firm IDs that enter the procedure.

**Person and firm IDs.** Persons are identified by their CPR-number. All individuals residing legally in Denmark are registered in the CPR register (Centrale Person Register) with a unique CPR number. The CPR-number is the sole identifier of an individual in relation to the state. Person data is recorded under an anonymized CPR-number labeled `pnr`.

There are two main ways of identifying business entities. The Central Business Register (CVR), established in October 1999, contains primary data on all businesses with economic activity in Denmark, regardless of economic and organizational structure. CVR covers both public and private businesses. The SE-number identifies a business in Stamregistret for Erhvervsdrivende (the SE-register), a register established in 1985. The main function of the SE-register is to identify businesses vis-a-vis the tax authorities when settling value added tax (VAT), income tax payments, or tax payments for self-employed persons, and pension contributions. CVR- and SE-numbers are recorded under the variable names `cvrnr` and `senr`, respectively. Although the relationship between CVR-numbers and SE-numbers is somewhat complicated by the fact that a business entity may split up into multiple legal business entities, an SE-number is associated with one and only one CVR-number at a given point in time. We are able to map CVR-numbers and SE-numbers using a correspondence table provided by Statistics Denmark.

As it turns out, employing firms in the labor market spell data are identified by the firm ID `spell_firm_id`, a hybrid firm identifier specific to the labor market spell data, which corresponds very closely to the CVR-number. As we describe below, we are able to map `spell_firm_id` to `cvrnr` using IDA-N data.

**The labor market spell data.** The unit of observation in the labor market spell data is a person-spell-year, where a spell is a job with a particular employer, and the employing firm is identified by `spell_firm_id` as described above. That is, a labor market spell that stretches across three calendar years (i.e. stretches across three January 1st) is represented by three observations in the labor market spell data.

**Merging IDA-P onto the labor market spell data.** We first merge the employment spell data and IDA-P. IDA-P contains annual background information from public registers on all individuals aged 15-74 residing legally in Denmark on the 31st of December. The unit of observation in IDA-P is a person-year. We merge on person identifier (variable `pnr`) and year (variable `aar`). 98.64% of the employment spell observations are matched to an IDA-P observation. We retain only person-years that are found in both data sources.

**Merging IDA-S onto the labor market spell data.** IDA-S contains background information from public records on all physical workplaces in Denmark, excluding the so-called fictitious workplaces briefly described in the main text. The unit of observation in the raw IDA-S panel is a workplace-year. Our analysis is conducted at the firm-level, and a firm may consist of several workplaces. Because IDA-S contains both workplace and firm identifiers, we can aggregate IDA-S to firm-years, letting a firm inherit industry affiliation and public sector status from its largest workplace.<sup>3</sup>

We start by mapping the firm identifier of employment spell data (variable `spell_firm_id`) into the firm identifier of the IDA database (`cvrnr`). The mapping between `spell_firm_id` and `cvrnr` is established using IDA-N, yet another IDA dataset. IDA-N contains data on all primary employment relationships ongoing on November 28th in a given year. The unit of observation in IDA-N is a person-workplace-year, but it also contains information on both the IDA-S firm identifier `cvrnr` and the spell data firm identifier (`spell_firm_id`), allowing us to construct a correspondence table with `spell_firm_id`, `cvrnr`, and `aar`. We use this table to map `spell_firm_id` into `cvrnr` in the labor market data,<sup>4</sup> and then we subsequently merge IDA-S onto the employment spells on firm identifier `cvrnr` and year `aar`. 96.10% of the employment spell observations are matched with an IDAS observation. We retain all employment spell data, whether matched to IDA-S or not.

**Merging VAT data onto the labor market spell data.** As described in the main text, value added data comes from two data sources, MOMS and MOMM. MOMS covers 1995-1999 and contains annual data on firms' sales and purchases, whereas MOMM starts in January, 2000, and provides the same information, but recorded on a monthly frequency.<sup>5</sup>

---

<sup>3</sup>The size of the workplace is the number of workers with primary employment at the workplace on the 28th of November, as measured in IDA-S

<sup>4</sup>There is a few occurrences where the mapping between `spell_firm_id` and `cvrnr` is 1-to-many, in which case we arbitrarily pick `cvrnr` as `max(cvrnr)`.

<sup>5</sup>The VAT declaration frequency depends on the firm's annual turnover. Firms register VAT information at the tax authorities monthly if their turnover exceeds DKK 15 mill., quarterly if their turnover is between DKK 1 mill. and DKK 15 mill., and every six months if turnover is below DKK 1 mill. The reporting frequency for a given firm is also provided in the MOMM data.



The MOMS data is recorded under the `senr`-firm ID, whereas the labor market spell data (and MOMM) is recorded under the firm ID `cvrnr`. As a first step in the merging of VAT data to the labor market spell data, we map `senr` in MOMS to `cvrnr`. This is done using a correspondence table provided by Statistics Denmark. Then, using MOMM data up until December, 2005, we aggregate the monthly data to annual frequencies. Stacking the MOMS and MOMM data now yields a value added panel covering the period 1995-2005, containing essentially all firms with economic activity in Denmark over the period. The unit of observation in this combined MOMS and MOMM data is a firm-year, with the firm ID given by `cvrnr`.

Merging the annual value added panel onto the labor market spell data, we match 90.37% of the employment panel firm-months to value added data.<sup>6</sup> We retain all employment spell data, whether matched to the VAT-data or not. This completes the merging procedure. As mentioned in the main text, the resulting 1990-2005 matched employer-employee panel contains 84,474,045 spell observations on 4,447,401 individuals, 428,448 firms, 57,568,393 job spells and 26,905,652 non-employment spells.

---

<sup>6</sup>The share of observations in the value added panel matched to the employment panel is much lower due to many firms in the value added data being small firms with no employees.

### A.3 Estimating $T'(\cdot)$ and $T(\cdot)$

Consider the 1999 - 2003 regime. Given that this regime has a progressive structure with three brackets (see Table 2.5), we impose the following flexible parametric function for marginal tax rates  $T'(\cdot)$ ,

$$T'(w) = \frac{\alpha_1}{1 + \exp(-\alpha_2(w - \alpha_3))} + \frac{\alpha_4}{1 + \exp(-\alpha_5(w - \alpha_6))} + \alpha_7. \quad (\text{A.52})$$

The marginal tax function  $T'(\cdot)$  is composed of two logistic functions “splined” together.

The data points we use to compute the parameters  $\alpha_1, \alpha_2, \dots, \alpha_7$  consist of 100 hourly wage distribution percentiles and 100 median marginal tax rates for each of the 5 years. Let  $(w_{tk}^{(50)}, [T']_{tk}^{(50)})$  for  $t = 1999, 2000, \dots, 2003$  and  $k = 1, 2, \dots, 100$  denote these data points. The parameters  $\alpha_1, \alpha_2, \dots, \alpha_7$  are estimated by fitting  $T'(w_i)$  to  $t_i$ , where  $T'(\cdot)$  is given by (A.52). That is,

$$\min_{\alpha_1, \alpha_2, \dots, \alpha_7} \sum_{t=1999}^{2003} \sum_{k=1}^{100} \left[ \frac{\alpha_1}{1 + \exp(-\alpha_2(w_{tk}^{(50)} - \alpha_3))} + \frac{\alpha_4}{1 + \exp(-\alpha_5(w_{tk}^{(50)} - \alpha_6))} + \alpha_7 - [T']_{tk}^{(50)} \right]^2. \quad (\text{A.53})$$

Let  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_7$  be the solution to the minimization problem (A.53).

Integrating (A.52) produces the following expression for the average tax function  $T(\cdot)$ ,

$$T(w) = \frac{\alpha_1}{\alpha_2} \log(e^{\alpha_2 \alpha_3} + e^{\alpha_2 w}) + \frac{\alpha_4}{\alpha_5} \log(e^{\alpha_5 \alpha_6} + e^{\alpha_5 w}) + \alpha_7 w + \alpha_0, \quad (\text{A.54})$$

where  $\alpha_0$  is the constant of integration. Parameter estimates for  $\alpha_1, \alpha_2, \dots, \alpha_7$  are available, and we estimate  $\alpha_0$  in the same way we estimated  $\alpha_1, \alpha_2, \dots, \alpha_7$ . That is, we calculate median tax liability  $T_{tk}^{(50)}$ , measured in the same unit as  $w_{tk}$ , for each of the 100 wage percentiles and for each of the years 1999 - 2003, and fit (A.54) with  $\alpha_1 = \hat{\alpha}_1, \alpha_2 = \hat{\alpha}_2, \dots, \alpha_7 = \hat{\alpha}_7$  to  $(w_{tk}, T_{tk}^{(50)})$  with  $t = 1999, 2000, \dots, 2003$  and  $k = 1, 2, \dots, 100$ :

$$\min_{\alpha_0} \sum_{t=1999}^{2003} \sum_{k=1}^{100} \left[ \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \log(e^{\hat{\alpha}_2 \hat{\alpha}_3} + e^{\hat{\alpha}_2 w_{tk}}) + \frac{\hat{\alpha}_4}{\hat{\alpha}_5} \log(e^{\hat{\alpha}_5 \hat{\alpha}_6} + e^{\hat{\alpha}_5 w_{tk}}) + \hat{\alpha}_7 w_{tk} + \alpha_0 - T_{tk}^{(50)} \right]^2. \quad (\text{A.55})$$

The marginal and average tax functions pertaining to the other tax regimes under consideration in this paper, i.e. the 1990-1993, the 1994-1998, and the 2004-2005 tax regimes, are estimated by analogous procedures, sometimes adapted to account for specific features of a tax regime. For example, for the 1990-1993 tax regime, and for the years 1994 and 1995, we build the marginal tax functions from three, rather than two, logistic functions.

## A.4 Numerical implementation

### A.4.1 Solving the model

#### A.4.1.1 Discretization of $H(a)$ and $\Gamma_0(p)$

Our algorithm is based on discrete values of worker ability  $a$  and firm productivity  $p$ , whereas the theoretical model takes the distributions  $H(a)$  and  $\Gamma_0(p)$  to be continuously differentiable. We apply Proposition B of Kennan (2006) to discretize the support of  $H(\cdot)$ . Let  $N_a$  and  $N_p$  be the numbers of discrete points in the support of worker ability  $a$  and firm productivity  $p$ , respectively.<sup>7</sup> For  $i = 1, \dots, N_a$ ,  $a_i = H^{-1}([2i - 1]/2N_a)$ , where  $H(\cdot)$  is the (continuous) population distribution of  $a$ .<sup>8</sup> We then calculate a probability mass function (PMF)  $\hat{h}(a_i)$  and a corresponding cumulative distribution function (CDF)  $\hat{H}(a_i)$  as  $\hat{h}(a_i) = 1/N_a$  and  $\hat{H}(a_i) = i/N_a$ , introducing the “^”-notation to distinguish the continuous population distributions and their discrete approximations. The same procedure applies for  $p$ , where we compute  $\hat{\gamma}_0(p_i) = 1/N_p$  and  $\hat{\Gamma}_0(p_i) = i/N_p$ .

#### A.4.1.2 Finding $\lambda(a)$ , $v(a|p)$ , and $w(a|p)$

This is the main part of numerical implementation. Given  $\hat{H}(a)$  and  $\hat{\Gamma}_0(p)$ , the structure of the algorithm is as follows.

0. Guess  $\lambda_0(a)$ ,  $s(w|a)$ ,  $v_0(a|p)$ , and  $w_0(a|p)$
1. Calculate  $s_1(w|a)$ ,  $w_1(a|p)$ ,  $v_1(a|p)$ , and  $\lambda_1(a)$
2. If  $s_1(w|a)$ ,  $w_1(a|p)$ ,  $v_1(a|p)$ , and  $\lambda_1(a)$  differ from  $s_0(w|a)$ ,  $w_0(a|p)$ ,  $v_0(a|p)$ , and  $\lambda_0(a)$ , take  $s_0(w|a) = s_1(w|a)$ ,  $w_0(a|p) = w_1(a|p)$ ,  $v_0(a|p) = v_1(a|p)$ , and  $\lambda_0(a) = \lambda_1(a)$  and go to 1. If not, take  $s(w|a) = s_1(w|a)$ ,  $w(a|p) = w_1(a|p)$ ,  $v(a|p) = v_1(a|p)$ , and  $\lambda(a) = \lambda_1(a)$  and terminate the algorithm.

We explain calculations of step 1 in detail below.

**Calculation of  $F(w(a|p)|a)$ .** The first step is to calculate  $F(w(a|p)|a)$ . For the current guesses of  $\lambda(a)$ ,  $v(a|p)$ , and  $w(a|p)$ , from (2.28), the discrete approximation of  $F(w(a_i|p_j)|a_i)$  is given by

$$\hat{F}(w(a_i|p_j)|a_i) = \frac{\sum_{z=p_0}^{p_j} v(a_i|z) \hat{\gamma}_0(z)}{\sum_{z=p_0}^1 v(a_i|z) \hat{\gamma}_0(z)}.$$

The associated PMF  $\hat{f}(w(a_i|p_j)|a_i)$  is calculated as  $\hat{F}(w(a_i|p_j)|a_i) - \hat{F}(w(a_i|p_{j-1})|a_i)$ .

<sup>7</sup>In the empirical implementation we take  $N_a = 100$  and  $N_p = 1,000$ .

<sup>8</sup>Recall that  $H(\cdot) = \mathcal{U}[0, 1]$  in our empirical implementation.

**Calculation of  $s(w|a)$ .** For the current guesses of  $\lambda(a)$ ,  $v(a|p)$ , and  $w(a|p)$ , we employ the following algorithm to compute  $s(w|a)$ .

0. Guess  $s(w_j|a_i) = s_0$ .
1. Calculate  $n^0(a_i)$  from (2.14) and  $\hat{G}(w_j|a_i)$  from (2.16).
2. Calculate  $s_1$  as the RHS of (2.13), replacing  $F(w|a)$  with  $\hat{F}(w_j|a_i)$ , and approximating the integral as detailed further below.
3. If  $s_1 \neq s_0$ , take  $s_0 = s_1$  and go to 1. If  $s_1 = s_0$ , take  $s(w_j|a_i) = s_1$  and terminate the algorithm.

Step 2 involves solving for  $s(w_j|a_i)$  as the fixed point of (2.13). We approximate the integral in (2.13) as follows:

$$\begin{aligned}
\int_{w_{ij}}^{\bar{w}(a_i)} \frac{u'(c(a_i))[1 - T'(x)]\bar{F}(x|a_i)}{\rho + \delta(a_i) + \lambda\mu + \lambda s(x|a_i)\bar{F}(x|a_i)} dx &= \int_{w_{ij}}^{w_{iN_p}} \phi_i(x) dx \\
&= \int_{w_{ij}}^{w_{iN_p}} \phi_i(x) \frac{1}{f(x|a_i)} dF(x|a_i) \\
&= \sum_{k=j}^{N_p} \phi_i(w_{ik}) \frac{1}{f(w_{ik}|a_i)} \hat{f}(w_{ik}|a_i) \\
&= \sum_{k=j}^{N_p-1} \phi_i(w_{ik}) \frac{1}{f(w_{ik}|a_i)} \hat{f}(w_{ik}|a_i) \\
&\approx \sum_{k=j}^{N_p-1} \phi_i(w_{ik}) \frac{\hat{f}(w_{ik}|a_i)}{\hat{f}(w_{i,k+1}|a_i)} (w_{i,k+1} - w_{ik}),
\end{aligned}$$

where  $\phi_i(x) \doteq \frac{u'(c(a_i))[1 - T'(x)]\bar{F}(x|a_i)}{\rho + \delta(a_i) + \lambda\mu + \lambda s(x|a_i)\bar{F}(x|a_i)}$ . For the second equality above, we use the definition of a PDF, i.e.,  $dF(x|a_i) = f(x|a_i)dx$ , while the third equality comes from the definition of expectation in discrete  $w$ . The fourth equality is from the fact that  $\phi_i(w_{iN_p}) = 0$  for all  $i$ . We finally use numerical differentiation to approximate  $f(w_{ik}|a_i) \approx \frac{F(w_{i,k+1}|a_i) - F(w_{ik}|a_i)}{w_{i,k+1} - w_{ik}} = \frac{\hat{f}(w_{i,k+1}|a_i)}{w_{i,k+1} - w_{ik}}$  for  $k \leq N_p - 1$ . Note that  $s(w_{iN_p}|a_i) = s(\bar{w}(a_i)|a_i) = 0$  for all  $i$ .<sup>9</sup>

---

<sup>9</sup>We compared  $s(w_{ij}|a_i)$  calculated from the method explained here with the one calculated from value function iteration, and confirmed that two methods gave the same results with negligible approximation errors. Due to computational time, our method is preferred to value function iteration.

#### A.4.1.3 Calculate $w(a|p)$

From (2.24),  $w(a_i|p_j)$  is calculated as

$$w(a_i|p_j) = y(a_i, p_j) - \frac{\tilde{\pi}(a_i|p_j)}{\tilde{\ell}(a_i|w_{i,j})}.$$

$\tilde{\ell}(a_i|w_{i,j})$  is calculated from (2.22). We approximate  $\tilde{\pi}(a_i|p_j)$  by the second order Taylor series sequentially:

$$\tilde{\pi}(a_i|p_{j+1}) \approx \tilde{\pi}(a_i|p_j) + \frac{\partial \tilde{\pi}(a_i|p_j)}{\partial p}(p_{j+1} - p_j) + \frac{1}{2} \cdot \frac{\partial^2 \tilde{\pi}(a_i|p_j)}{\partial p^2}(p_{j+1} - p_j)^2. \quad (\text{A.56})$$

By the Envelope Theorem,

$$\frac{\partial \tilde{\pi}(a_i|p_j)}{\partial p} = \frac{\partial y(a_i, p_j)}{\partial p} \tilde{\ell}(a_i|w_{i,j}).$$

By taking the second derivative and applying numerical differentiation,

$$\begin{aligned} \frac{\partial^2 \tilde{\pi}(a_i|p_j)}{\partial p^2} &= \frac{\partial^2 y(a_i, p_j)}{\partial p^2} \tilde{\ell}(a_i|w_{i,j}) + \frac{\partial y(a_i, p_j)}{\partial p} \cdot \frac{\partial \tilde{\ell}(a_i|w_{i,j})}{\partial w} \cdot \frac{\partial w_{i,j}}{\partial p} \\ &\approx \frac{\partial^2 y(a_i, p_j)}{\partial p^2} \tilde{\ell}(a_i|w_{i,j}) + \frac{\partial y(a_i, p_j)}{\partial p} \cdot \frac{\tilde{\ell}(a_i|w_{i,j+1}) - \tilde{\ell}(a_i|w_{i,j})}{w_{i,j+1} - w_{i,j}} \cdot \frac{w_{i,j+1} - w_{i,j}}{p_{j+1} - p_j}. \end{aligned}$$

Thus, by substituting these partial derivatives, we can calculate (A.56) as

$$\tilde{\pi}(a_i|p_{j+1}) \approx \tilde{\pi}(a_i|p_j) + \frac{\partial y(a_i, p_j)}{\partial p} \cdot \frac{\tilde{\ell}(a_i|w_{i,j}) + \tilde{\ell}(a_i|w_{i,j+1})}{2} (p_{j+1} - p_j) + \frac{\partial^2 y(a_i, p_j)}{\partial p^2} \cdot \frac{\tilde{\ell}(a_i|w_{i,j})}{2} (p_{j+1} - p_j)^2.$$

The initial condition is given by  $\tilde{\pi}(a_i|p_0) = 0$  for all  $a_i$  from  $\pi(p_0) = 0$ .

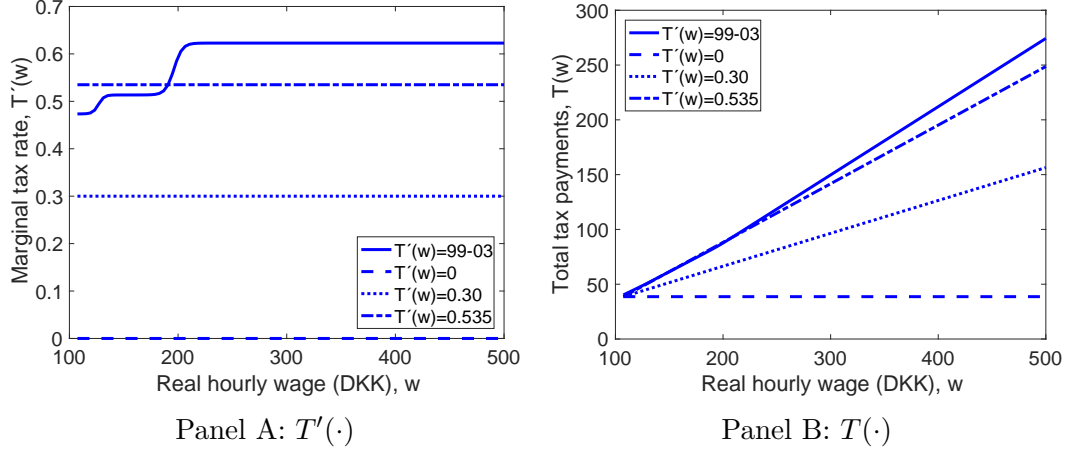
#### A.4.1.4 Calculate $v(a|p)$

$M_0$  is given by (2.27), which gives  $V(a) = M_0 \int_{p_0}^1 v(p) d\Gamma_0(p)$ . We calculate  $v(a|p)$  from (2.23).

#### A.4.1.5 Calculate $\lambda(a)$

$S(a)$  is given by (2.29), which gives  $\lambda(a) = (V(a)/S(a))^{1-\eta}$ .

Figure E.1: Counterfactual linear tax systems



## A.5 Allocative distortions and tax progressivity

Our evaluations of the 1990, 1994, 1999, and 2004 Danish tax reforms as well as the analysis of the elasticity of taxable income suggest that income taxations induce quantitative relevant distortions in the equilibrium allocation of labor. The Danish income tax schedule is characterized by a high degree of progressivity, see Figures 2.7, and in this section we consider a number of counter-factual reforms to quantify the impact of income tax progressivity on the allocation of labor.

Specifically, we consider three counterfactual linear tax regimes, with  $T'(w) = 0.54$ ,  $T'(w) = 0.30$ , and  $T'(w) = 0$ . The  $T'(w) = 0.54$  is the mechanically revenue neutral linear tax regime, that is, in the absence of any behavioral changes, replacing the actual 1999-2003 income tax system with a the linear tax function  $T'(w) = 0.54$  would leave the government budget unchanged. The  $T'(w) = 0.30$  tax regime is intended as a brute approximation of the US income tax system, and the extreme counterfactual  $T'(w) = 0$  tax system is included primarily for illustration. When we impose each of the counterfactual tax systems, we always impose that  $b - T(b)$  stays constant. The counterfactual tax regimes are rendered graphically in Figure E.1 along with the actual 1999-2003 income tax system.

Table E.1 present the percentage changes in nonemployment, labor income and the government budget for each of the three counterfactual tax regimes. As usual, we report both partial and full equilibrium responses, and also report the compensated responses where we keep the marginal utility of consumption constant. We stress again, the we do not balance the government budget in these cases, an issue that may be particularly egregious in the context of large counterfactual reforms where the direct effect on the

government budget may be rather large, and one might suspect that income effects arising from lump sum transfers balancing the budget may be quantitatively important.<sup>10</sup>

The mechanically revenue neutral linear tax  $T'(w) = 0.54$  increases nonemployment, and reduced expected labor income in both full and partial equilibrium almost everywhere in the ability distribution, except at the very top. In the aggregate, the  $T'(w) = 0.54$  regime therefore induces an increase in the nonemployment rate, a decrease in aggregate labor income, as well as a small reduction in the government budget stemming from changes in the behavior of workers and firm.

Moving on to the  $T'(w) = 0.30$  regime, we see relatively large reduction in the nonemployment rate, especially in partial equilibrium, and fairly large increases in labor income, across the ability distribution. This implies that the aggregate nonemployment rate is reduced by almost 10 percent in full equilibrium, while aggregate labor income increases by 5 percent. The impact on the government budget is very large, with the full equilibrium effect being a 24 percent drop in government revenue. We caution that if we were to make workers absorb this reduction in the budget through lump-sum taxation, it would likely induce rather large income effects in job search behavior. These could alter the conclusion drawn regarding the effect of going to a  $T'(w) = 0.30$  regime on nonemployment and labor income. The same caveats apply to the interpretation of the last counterfactual exercise presented in Table E.1, where we impose the extreme tax regime  $T'(w) = 0$ , while keeping  $b - T(b)$  constant. This reform has the same qualitative predictions as discussed above for the case of  $T'(w) = 0.30$  regime, but quantitatively, the impact is of course large, and that includes a very large and negative impact on the government budget.

---

<sup>10</sup>As mentioned above, we are currently working on amending our computation algorithms to impose balanced government budgets, but are at present unable to present empirical results for this case.

Table E.1: The impact of income tax progressivity on nonemployment, labor income and the government budget

			$\frac{\Delta n^0}{n^0} \times 100\%$			$\frac{\Delta LI}{LI} \times 100\%$			$\frac{\Delta B}{B} \times 100\%$		
Comparison tax regime	Ability	Total		Subst		Total		Subst		Total	
		PE	FE	PE	FE	PE	FE	PE	FE	PE	FE
$T'(w) = 0.54$	All	2.9	5.7	2.3	5.5	-0.4	-0.6	-0.3	-0.4	-1.0	-2.1
	0.25	8.4	14.4	7.8	15.2	-2.6	-5.6	-2.4	-5.9	-5.8	-14.5
	0.50	4.5	5.6	4.9	6.9	-0.5	-1.2	-0.5	-1.5	2.1	1.2
	0.75	0.9	2.4	-0.1	1.1	0.1	1.4	0.1	1.7	-1.2	0.1
$T'(w) = 0.30$	All	-10.7	-5.0	-18.3	-9.8	1.6	3.1	2.4	5.0	-24.2	-24.1
	0.25	-13.9	-10.2	-17.1	-9.7	4.2	4.5	5.2	4.4	4.7	2.2
	0.50	-12.8	-5.3	-20.4	-9.9	1.4	2.0	2.3	3.4	-15.4	-16.1
	0.75	-12.0	-3.3	-26.2	-14.6	1.0	3.2	1.9	5.9	-24.5	-24.2
$T'(w) = 0$	All	-19.6	-11.8	-32.3	-20.5	2.9	5.7	4.4	9.2	-56.8	-57.4
	0.25	-29.0	-24.6	-34.4	-25.3	8.8	11.0	10.5	11.4	2.3	-2.9
	0.50	-23.8	-12.1	-37.2	-21.6	2.7	4.1	4.2	7.2	-40.6	-43.0
	0.75	-20.1	-6.5	-41.6	-24.2	1.3	3.5	3.1	9.1	-57.0	-58.6



# Bibliography

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies* 72(1), 1–19.
- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association* 105(490), 493–505.
- Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science* 59(2), 495–510.
- Abadie, A. and J. Gardeazabal (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* 93(1), 113–132.
- Abowd, J. M., R. H. Creedy, and F. Kramarz (2002, March). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.
- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67(2), 251–333.
- Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. MIT Press.
- Adam, S., T. Besley, R. Blundell, S. Bond, R. Chote, M. G. P. Johnson, G. Myles, and J. M. Poterba (2011). *Dimension of Tax Design*. Oxford University Press.
- Altonji, J. G. and C. H. Paxson (1992). Labor Supply, Hours Constraints, and Job Mobility. *Journal of Human Resources* 27(2), 256–278.
- Altonji, J. G. and L. M. Segal (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics* 14, 353–366.
- Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: an empiricist’s companion*. Princeton University Press.

- Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *Review of Economics and Statistics* 60(1), 47–57.
- Auten, G. and R. Carroll (1999). The Effect of Income Taxes on Household Income. *Review of Economics and Statistics* 81(4), 681–693.
- Bagger, J. and R. Lentz (2015). An empirical model of wage dispersion with sorting.
- Bagger, J., E. R. Moen, and R. M. Vejlin (2017). Optimal taxation with on-the-job search.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-In-Differences Estimates? *Quarterly Journal of Economics* 119(1), 249–275.
- Blomquist, S. and H. Selin (2010). Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates. *Journal of Public Economics* 94(11-12), 878–889.
- Blundell, R., M. Brewer, and M. Francesconi (2008). Job Changes and Hours Changes: Understanding the Path of Labor Supply Adjustment. *Journal of Labor Economics* 26(3), 421–453.
- Blundell, R., M. Costa Dias, C. Meghir, and J. van Reenen (2004). Evaluating the Employment Impact of a Mandatory Job Search Program. *Journal of the European Economic Association* 2(4), 569–606.
- Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. *Handbook of Labor Economics* 3, 1559–1695.
- Blundell, R. and T. MaCurdy (1999). *Labour Supply: A Review of Alternative Approaches*, Volume 3 of *Handbook of Labor Economics*, Chapter 27. Elsevier Inc.
- Blundell, R., T. MaCurdy, and C. Meghir (2007). Labor Supply Models: Unobserved Heterogeneity, Nonparticipation and Dynamics. *Handbook of Econometrics* 6, 4667–4775.
- Blundell, R. and A. Shephard (2012). Employment, hours of work and the optimal taxation of low income families. *Review of Economic Studies* 79(2), 481–510.
- Bontemps, C., J.-M. Robin, and G. J. van den Berg (1999). An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms. *International Economic Review* 40(4), 1039–1074.

- Bontemps, C., J.-M. Robin, and G. J. van den Berg (2000). Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review* 41(2), 305–358.
- Breda, T., L. Haywood, and H. Wang (2017). Labor market responses to payroll tax reductions.
- Burdett, K. (1978). A theory of employee job search and quit rates. *American Economic Review* 68(1), 212–220.
- Burdett, K. and M. Coles (2003). Equilibrium wage tenure contracts. *Econometrica* 71(5), 1377–1404.
- Burdett, K. and M. Coles (2010). Wage/tenure contracts with heterogeneous firms. *Journal of Economic Theory* 145(4), 1408–1435.
- Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica* 51(4).
- Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review* 39(2), 257–273.
- Burns, S. K. and J. P. Ziliak (2017). Identifying the Elasticity of Taxable Income. *Economic Journal* 127(600), 297–329.
- Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica* 74(2), 323–364.
- Chetty, R. (2009a). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy* 1(2), 31–52.
- Chetty, R. (2009b). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics* 1(1), 451–487.
- Chetty, R., J. Friedman, T. Olsen, and L. Pistaferri (2011a). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics* 126, 749–804.
- Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011b). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics* 126(2), 749–804.

- Christensen, B. J., R. Lentz, D. T. Mortensen, G. R. Neumann, and A. Werwatz (2005). On-the-Job Search and the Wage Distribution. *Journal of Labor Economics* 23(1), 31–58.
- Coles, M. (2001). Equilibrium wage dispersion, firm size, and growth. *Review of Economic Dynamics* 4, 159–187.
- Diamond, P. A. (1971). A theory of price adjustment. *Journal of Economic Theory* 3, 156–168.
- Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy* 103(3), 551–572.
- Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics* 81(4).
- Fuest, C., A. Peichl, and S. Siegloch (2018). Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany. *American Economic Review* 108(2), 393–418.
- Gentry, W. M. and R. G. Hubbard (2004a). The effects of progressive income taxation on job turnover. *Journal of Public Economics* 88, 2301–2322.
- Gentry, W. M. and R. G. Hubbard (2004b). The Effects of Progressive Income Taxation on Job Turnover. *Journal of Public Economics* 88(11), 2301–2322.
- Giertz, S. H. (2010). The Elasticity of Taxable Income during the 1990s: New Estimates and Sensitivity Analyses. *Southern Economic Journal* 77(2), 406–433.
- Goolsbee, A. (2000). What Happens When You Tax the Rich? Evidence from Executive Compensation. *Journal of Political Economy* 108(2), 352–378.
- Gorry, A., K. A. Hassett, R. G. Hubbard, and A. Mathur (2017). The response of deferred executive compensation to changes in tax rates. *Journal of Public Economics* 151, 28–40.
- Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Gruber, J. and E. Saez (2002). The elasticity of taxable income: Evidence and implications. *Journal of Public Economics* 84(1), 1–32.
- Güvenen, F., B. Kuruscu, and S. Ozkan (2014). Taxation of human capital and wage inequality: A cross-country analysis. *Review of Economic Studies* 81(2), 818–850.

- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis* 20(1), 25–46.
- Hainmueller, J. and Y. Xu (2013). Ebalance: a Stata package for entropy balancing. *Journal of Statistical Software* 54(7), 1–18.
- Hall, B. J. and J. B. Liebman (2000). The Taxation of Executive Compensation. *Tax Policy and the Economy* 14, 1–44.
- Hall, R. E. and A. B. Krueger (2012). Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search. *American Economic Journal: Macroeconomics* 4(4), 56–67.
- Heathcote, J., K. Storesletten, and G. Violante (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics* 132(4), 1693–1754.
- Heckman, J. (1993). What has been learned about labor supply in the past twenty years? *American Economic Review* 83(2), 116–121.
- Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64(4), 605–654.
- Heckman, J. J., L. Lochner, and C. Taber (1998). Tax Policy and Human-Capital Formation. *American Economic Review* 88(2), 293–297.
- Heckman, J. J., L. Lochner, and C. Taber (1999). Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy. *Fiscal Studies* 20(1), 25–40.
- Heim, B. T. (2009). The Effect of Recent Tax Changes on Taxable Income: Evidence from a New Panel of Tax Returns. *Journal of Policy Analysis and Management* 28(1), 147–163.
- Hendren, N. (2016). The Policy Elasticity. *Tax Policy and the Economy* 30(1), 51–89.
- Holmlund, B. and M. Soderstrom (2011). Estimating Dynamic Income Responses to Tax Reform. *B.E. Journal of Economic Analysis & Policy* 11(1).
- Hosseini, R. and A. Shourideh (2017). Retirement financing: An optimal reform approach.
- Imbens, G. W. (2015). Matching Methods in Practice: Three Examples. *Journal of Human Resources* 50(2), 373–419.

- Jakobsen, K., K. Jakobsen, H. Kleven, and G. Zucman (2018). Wealth taxation and wealth accumulation: Theory and evidence from denmark. Technical report, NBER Working Paper 24371.
- Jolivet, G., F. Postel-Vinay, and J.-M. Robin (2006). The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US. *European Economic Review* 50(4), 877–907.
- Keane, M. P. (2011). Labor Supply and Taxes: A Survey. *Journal of Economic Literature* 49(4), 961–1075.
- Keane, M. P. (2015). Effects of Permanent and Transitory Tax Changes in a Life-Cycle Labor Supply Model With Human Capital. *International Economic Review* 56(2), 485–503.
- Keane, M. P. (2016). Life-cycle Labour Supply with Human Capital: Econometric and Behavioural Implications. *Economic Journal* 126(592), 546–577.
- Keane, M. P. and N. Wasi (2016). Labour Supply: The Roles of Human Capital and The Extensive Margin. *Economic Journal* 126(592), 578–617.
- Kennan, J. (2006). A note on discrete approximation of continuous distributions.
- Kleven, H. J. (2014). How can Scandinavians tax so much? *Journal of Economic Perspectives* 28(4), 77–98.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics* 8(1), 435–464.
- Kleven, H. J. (2018). Sufficient Statistics Revisited.
- Kleven, H. J. and E. A. Schultz (2014a). Estimating Taxable Income Responses Using Danish Tax Reforms. *American Economic Journal: Economic Policy* 6(4), 271–301.
- Kleven, H. J. and E. A. Schultz (2014b). Estimating taxable income responses using danish tax reforms. *American Economic Journal: Economic Policy* 6(4), 271–301.
- Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics* 89(11-12), 2093–2119.
- Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from Danish monthly payroll records. *American Economic Journal: Economic Policy* 8(3), 233–257.

- Kreiner, C. T., J. R. Munch, and H. J. Whitta-Jacobsen (2015). Taxation and the long run allocation of labor: Theory and Danish evidence. *Journal of Public Economics* 127, 74–86.
- le Maire, D. and B. Schjerning (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics* 107, 1–18.
- le Maire, D. and B. Schjerring (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics* 107, 1–18.
- Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics* 4(3), 165–224.
- Lentz, R. (2010, July). Sorting by search intensity. *Journal of Economic Theory* 145(4), 1436–1452.
- Lentz, R. (2015). Optimal employment contracts with hidden search.
- Lockwood, B. and A. Manning (1993). Wage setting and the tax system theory and evidence for the United Kingdom. *Journal of Public Economics* 52(1), 1–29.
- Lockwood, B., T. Slok, and T. Tranaes (2000). Progressive Taxation and Wage Setting: Some Evidence for Denmark. *Scandinavian Journal of Economics* 102(4), 707–723.
- Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy* 126(1), 313–346.
- Lund, C. G. and R. Vejlin (2015). Documenting and improving the hourly wage measure in the danish ida database. *Danish Journal of Economics*.
- Lund, C. G. and R. Vejlin (2016). Documenting and Improving the Hourly Wage Measure in the Danish IDA Database. *Danish Journal of Economics* 1, 1–35.
- Marcus, J. (2013). The effect of unemployment on the mental health of spouses - Evidence from plant closures in Germany. *Journal of Health Economics* 32(3), 546–558.
- Marcus, J. and T. Siedler (2015). Reducing binge drinking? The effect of a ban on late-night off-premise alcohol sales on alcohol-related hospital stays in Germany. *Journal of Public Economics* 123, 55–77.
- Meghir, C. and D. Phillips (2010). Labour Supply and Taxes. In *Dimensions of Tax Design: The Mirrlees Review*, pp. 202–274.

- Melitz, M. J. (2003). The impact of trade on intraindustry reallocations and aggregate industry productivity. *Econometrica* 71(6), 1695–1725.
- Mertens, K. and J. L. M. Olea (2018). Marginal Tax Rates and Income: New Time Series Evidence.
- Merz, M. (1995). Search in the labor market and the real business cycle. *Journal of Monetary Economics* 36(2), 269–300.
- Michau, J.-B. (2017). On the provision of insurance against search-induced wage fluctuations.
- Mirrlees, J., S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. G. P. Johnson, G. Myles, and J. M. Poterba (2011). *Tax by Design*. Oxford University Press.
- Mortensen, D. T. (2005). *Wage Dispersion: Why Are Similar Workers Paid Differently?* MIT Press.
- Piketty, T. and E. Saez (2013). *Optimal Labor Income Taxation*, Volume 5 of *Handbook of Public Economics*, Chapter 7, pp. 391–474. Elsevier Inc.
- Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal Taxation of Top Labour Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy* 6(1), 230–271.
- Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70(6), 2295–2350.
- Postel-Vinay, F. and J.-M. Robin (2004). To match or not to match? to match or not to match? optimal wage policy with endogenous worker search intensity. *Review of Economic Dynamics* 7, 297–330.
- Saez, E. (2003). The effect of marginal tax rates on income: A panel study of 'bracket creep'. *Journal of Public Economics* 87(5-6), 1231–1258.
- Saez, E., B. Schoefer, and D. Seim (2017). Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden.
- Saez, E., J. Slemrod, and S. H. Giertz (2012a). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economics Literature* 50(1), 3–50.



- Saez, E., J. Slemrod, and S. H. Giertz (2012b). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature* 50(1), 3–50.
- Shephard, A. (2017). Equilibrium search and tax credit reform. *International Economic Review* 58(4), 1047–1088.
- Shimer, R. (2010). *Labor Markets and Business Cycles*. Princeton University Press.
- Simonsen, M. and L. Skipper (2008). The Incidence and Intensity of Formal Lifelong Learning.
- Singleton, P. (2011). The Effect of Taxes on Taxable Earnings: Evidence From the 2001 and Related U.S. Federal Tax Acts. *National tax journal* 64(2), 323–351.
- Slemrod, J. (1992). Do Taxes Matter? Lessons from the 1980’s. *American Economic Review* 82(2), 250–256.
- Slemrod, J. (1995). Income creation or income shifting? behavioral responses to the tax reform act of 1986. *American Economic Review* 85(2), 175–180.
- Slemrod, J. (1996). High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response. In *Empirical Foundations of Household Taxation*, pp. 169–192.
- Slemrod, J. (1998). Methodological Issues in Measuring and Interpreting Taxable Income Elasticities. *National tax journal* 51(4), 773–788.
- Slemrod, J. and W. Kopczuk (2002). The optimal elasticity of taxable income. *Journal of Public Economics* 84(1), 91–112.
- Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde’s critique of nonexperimental estimators? *Journal of Econometrics* 125(1-2), 305–353.
- Sørensen, K. L. and R. Vejlin (2014). Return to Experience and Initial Wage Level: Do Low Wage Workers Catch Up? *Journal of Applied Econometrics* 29(6), 984–1006.
- Stevens, M. (2004). Wage-tenure contracts in a frictional labour market: Firms strategies for recruitment and retention. *Review of Economic Studies* 71(2), 535–551.
- Stinebrickner, R., T. R. Stinebrickner, and P. J. Sullivan (2017). Job Tasks, Time Allocation, and Wages.
- Suárez Serrato, J. C. and O. Zidar (2016). Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms. *American Economic Review* 106(9), 2582–2624.

- Taber, C. (2002). Tax Reform and Human Capital Accumulation: Evidence from an Empirical General Equilibrium Model of Skill Formation. *Advances in Economic Analysis and Policy* 2(1).
- Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *Quarterly Journal of Economics* 107(2), 439–479.
- Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics* 117, 90–103.
- Werning, I. (2007). Pareto efficient income taxation.
- Zhao, Q. and D. Percival (2017). Entropy Balancing is Doubly Robust. *Journal of Causal Inference* 5(1).